DOCUMENT RESUME

ED 064 440 UD 012 483

AUTHOR Davidson, Terrence N.
TITLE Youth in Transition, Volume IV. Evolution of a Strategy for Longitudinal Analysis of Survey Panel Data.
INSTITUTION Michigan Univ., Ann Arbor. Survey Research Center.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
PUB DATE 72
NOTE 130p.; Report written in partial fulfillment of the requirements for the Ph.D. degree, Univ. of Michigan, Ann Arbor, Mich.

EDRS PRICE MF-$0.65 HC-$6.58
DESCRIPTORS *Adolescents; *Data Analysis; High School Students; *Longitudinal Studies; Models; *National Surveys; Predictor Variables; Raw Scores; Social Environment; *Youth Problems

ABSTRACT

This report is an integral part of the publication series of the Youth in Transition study, a nationwide panel survey of adolescent boys, which attempts to discover and document how the contemporary social environments affect the development of young men during their high school years. Four waves of data were gathered from 2,213 boys comprising the sample, who were clustered into 87 different high schools throughout the country. Additionally, because of the special interest in the school environment, data were collected from the principals, counselors, and samples of teachers in the participating schools. The efforts reported in this study are based on the attempt to outline a practical procedure to be used in longitudinal analyses of Youth in Transition data, the major aim being to develop a strategy which can be applied to most, if not all, of the analyses to be performed. The report develops a "parallel prediction" model for longitudinal analysis, which makes separate use of each repetition of the criterion dimension; it is contended that the proposed strategy is widely applicable in studies employing panel designs. The proposed model was applied to a limited set of analyses of the Youth in Transition data. Early identification of subgroups was seen to have a facilitating effect in longitudinal analysis. (Author/RJ)

Youth in Transition

VOLUME IV

Evolution of a Strategy for Longitudinal Analysis of Survey Panel Data

TERRENCE N. DAVIDSON

SURVEY RESEARCH CENTER

ISR

INSTITUTE FOR SOCIAL RESEARCH
THE UNIVERSITY OF MICHIGAN
ANN ARBOR, MICHIGAN

ISR CODE #3404

The research reported herein was performed pursuant to a contract with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

Library of Congress Catalog Card Number: 67-66009

ISBN 87944-115-1

Printed by Litho Crafters, Inc., Ann Arbor, Michigan
Manufactured in the United States of America

First Printing 1972

PREFACE

This report is written in partial fulfillment of the requirements for the degree of Doctor of Philosophy in The University of Michigan. It is also an integral part of the publication series of the Youth in Transition study, a nationwide panel survey of adolescent boys. This study is attempting to discover and document how the contemporary social environments affect the development of young men during their high school years.

The first major data collection was fielded in the fall of 1966. The respondents were chosen so as to be representative of sophomore boys in public high schools in the United States that fall. The 2,213 boys comprising the sample were clustered into 87 different high schools throughout the country.[1]

Three additional waves of data have now been gathered from this sample of boys. Additionally, because of our special interest in the school environment, data were collected from the principals, counselors, and (samples of) teachers in each of

[1]For a more detailed description of the sampling design, see Bachman, et al., 1967, pp. 21-24.

the participating schools. These "organizational" data were gathered at about the same time as our second wave of data from the boys who were then just completing their junior years.[2] The combination of four waves of boys' data and the organizational data provides a rich set of measures for longitudinal analyses.

The efforts reported herein are based on the author's attempt to outline a practical procedure to be used in longitudinal analyses of the Youth in Transition data. The major aim of these efforts is to develop a strategy which can be applied to most, if not all, of the longitudinal analyses to be performed. This attempt to generate a general-purpose procedure is in keeping with a need expressed in the study's first discussion of its analysis design:

> Because of the broad scope of the project, and especially because of its longitudinal design, the possibilities for data analysis are vast. It is therefore essential that we develop systems and procedures of analysis that give high priority to data integration and that we provide strategies for examining many substantive questions simultaneously (Bachman, et al., 1967, p. 81).

Another characteristic of these efforts should be noted at this point. A project as large as Youth in Transition is best viewed as something

---

[2] Of course, some of the original respondents had left school by this time. However, we continued to include both "dropouts" and "stayins" in subsequent data collections.

other than an hypotheses-testing venture. We have previously attempted to describe our philosophical approach as one of "theoretically-guided empiricism" (Bachman, et al., 1967, p. 17). In doing so, we do not mean to denigrate the use of the hypothetical-deductive model which underlies much of the methodology associated with inferential statistics. Rather, we mean to emphasize that we see our major contribution primarily in the inductive phase of theory development and not in the deductive or model-testing phase.

A colleague has described such efforts as attempts to "find variables that work" (Sonquist, 1969, pp. 83-95). He describes the major problem as "...one of determining which of the variables for which data have been collected are actually related to the phenomenon in question, and under what conditions and through which intervening processes, with appropriate controls for spuriousness" (Sonquist, 1970, p. 1). Thus, we will make frequent use of techniques appropriate to this "finding variables that work" mission. In doing so, our approach may be described as an effort in "discovering grounded theory" (Glaser and Strauss, pp. 1-18).

Finally, because of the project's major emphasis on exploring the impact of the school environment, the analyses to be reported will focus on the 1,374 young men who remained in the same high school during the first three data collections. Concentrating these analyses on this non-moving, non-dropout subset should not be taken as an indication of disinterest in the other young men in the sample. Quite the contrary; specific analysis plans have already been made to examine other subsets of the data, and their results will be reported in forthcoming monographs. However, we do not want to confound our initial analyses of environmental effects by including boys who were not exposed to the same school environment throughout their high school years.

Acknowledgements

A project such as Youth in Transition could not be undertaken without a staff of dedicated and competent people. I am pleased to acknowledge the vital support they have provided me throughout the study as well as in the production of this volume. In addition, I am grateful to the Institute for Social Research for providing both a stimulating environment and the technical assistance to implement a research plan as complex as ours. I am especially indebted to the Computer Support Group of the Survey Research Center, headed by Judith Rattenbury, and to the Computer Services Facility, headed by Duane Thomas; without their talents and cooperation, I would not now be writing.

Among the many colleagues who afforded me a receptive audience as plans for this effort began to take shape, I am especially indebted to Jerald

Bachman, both for his wise counsel and for providing my salary, and to John Bonquist, who always sent me away with a better question than the one I brought him. Additionally, I am deeply appreciative of the technical and moral support of my doctoral committee: Co-Chairmen Jerald Bachman and William Cave, Frank Andrews, and LaVerne Collet. I also gratefully acknowledge the help of Robert Kahn who kindly provided me with several fruitful suggestions and who served as an ex-officio committee member.

I am indebted to Pam Deasy and Reggie Gerstman, both for their deciphering ability and for their complete mastery of a machine which has baffled me throughout my life--the typewriter. Their skills were essential in the production of this manuscript and they are most appreciated.

I fully acknowledge the assistance of all of these people. Whatever errors have escaped their eyes and remain yet in this manuscript are mine. I know that such errors are fewer because of the help I received from them.

Primary financial support for the Youth in Transition project was obtained from the Office of Education. Additional support for some phases of the research has been provided by the United States Department of Labor and the United States Department pf Defense. I am indebted to these three agencies for their vital support.

Finally, I am grateful to my parents for their encouragement and devotion to the pursuit of my goals, and for their patience in answering the question, "Is Terry still in school?"; to Joseph Johnston, Jr., who knew long before I did that this day would come for me and whose professional collaboration and personal friendship I am privileged to enjoy; to Ned Flanders for affording me the opportunity to return to school and for providing a stimulating group of fellow students who truly made studying enjoyable; and especially to my fiancee, Diane Knapp, a constant source of inspiration who provided the real motivation for me to receive my PhD before she receives her Mrs.

Youth in Transition Project Staff (Past and Present)

Allison Arscott	Rita Lamendella
Jerald Bachman	Judith Long
Joy Bingham	Martha Mednick
Lynn Bozoki	Haydee Navarro
Janet Bumpass	Roberta Niaki
Robert Cope	Guttorm Norstebo
Terrence Davidson	Patrick O'Malley
John French, Jr.	Karen Paige
Regina Gerstman	Janice Plotkin
Swayzer Green	Philip Rappaport
Penni Holt	Joel Raynor
Sally Iman	Willard Rodgers
Mary Jacobs	Claire Taylor
Jerome Johnston	Barbara Thomas
Lloyd Johnston	Elizabeth van Duinen
Robert Kahn	Patricia Veerkamp
Diane Davidson	Ilona Wirtanen

CONTENTS

TABLES

FIGURES

Chapter 1
INTRODUCTION TO THE IMPORTANT ISSUES

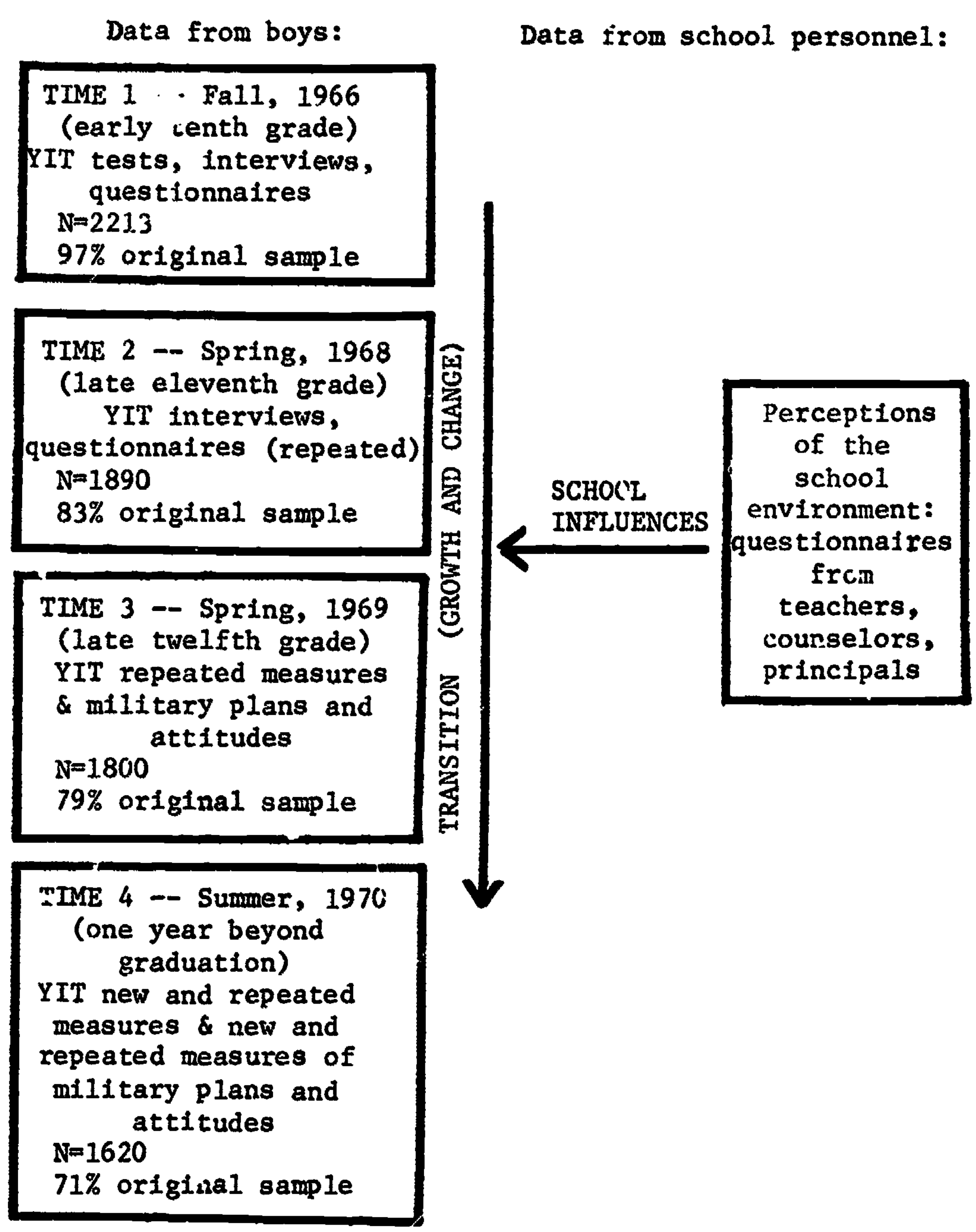
One of the important objectives of the Youth in Transition study is to examine some of the major changes taking place in adolescent boys during the high school years. Consequently, measures of many important dimensions were taken at each of the four data collections shown in Figure 1-1. These repeated dimensions may be placed into seven classes: motives, affective states, self-concept, values, attitudes, plans, and behaviors.[1] There are approximately 45 such dimensions which have been included in all four waves of data.

Now one of the prime areas of analysis is focused on attempting to explain how the immediate social environment affects the motives, values, plans, etc., of adolescent boys. Thus, we wish to consider to what extent these repeatedly measured dimensions may be predicted from characteristics of his home and family background,[2] school and peer group environment, and job environment. This

[1]For a more comprehensive description of the variables included in the study, see Bachman, et al., 1967, Chapter 4.

[2]The predictability of the initial measures is summarized in Bachman, 1970.

FIGURE 1-1
THE YOUTH IN TRANSITION STUDY
OVERVIEW OF RESEARCH DESIGN



brings into the picture several hundred measures of these important social environments to be considered as potentially important predictors of the repeatedly measured criterion dimensions.

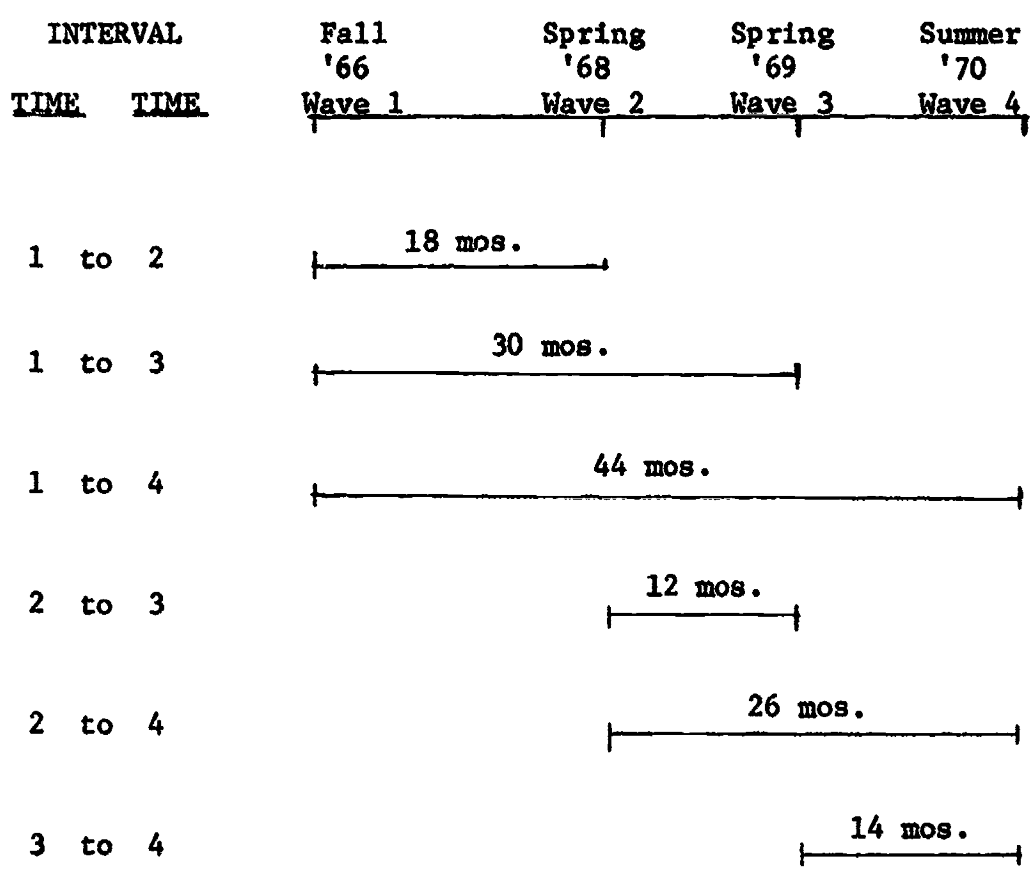
There is one additional complication which should also be mentioned. For each repeated criterion dimension, we will have four separate observations (one at each of the four data waves) available, each of considerable interest as a criterion variable in its own right. In addition, we can examine up to six kinds of changes by deriving measures which correspond to the intervals between each pair of measures. (See Figure 1-2.) In all then, we have potentially ten versions of each of the above-mentioned criteria, the four static or cross-sectional scores and the six dynamic or change scores.[3]

It is perhaps as obvious to the reader as it was to the study staff that a general purpose strategy is an absolute necessity when faced with the plethora of predictor and criterion variables to be analyzed. To put it another way, we have neither the time nor the desire to custom-build an analysis sequence for each criterion dimension (or for each predictor for that matter). Rather, what we seek is to develop a strategy which can be used to relate a large set of predictors to each of a large set of criteria. Much of what is

[3]As we shall see later, even this is an oversimplification because there are several competing methods of deriving each of the six change scores.

FIGURE 1-2

SIX POSSIBLE INTERVALS FOR ASSESSING CHANGE



presented in the following chapters may be accurately viewed as an attempt to develop such a strategy. But before turning to the results of these efforts, let us first examine briefly what is meant by the term "change."

What is Meant by "Change"?

Because there has been much confusion in the past about what is meant by "change," let us begin this chapter by attempting to clarify the author's use of this term. The most critical distinction arises when examining *individual differences* in change as opposed to *average* changes. As scientists, we are interested in what causes changes to occur in attitudes, values, aspirations, etc., even if some individuals change in one direction and others change in the opposite direction. In such a situation, we may observe no average change, but it may be most interesting to try to discover what variables are affecting these "equal and opposite" individual changes.

There are many situations in which a study of individual change would be rather fruitless. For example, consider a situation in which a set of uniform procedures have been designed to bring subjects to a specified terminal performance level. Now, suppose such procedures are administered to a group of subjects who are totally naive initially, and that those procedures are utilized until each subject attains the specified performance level.

In such a situation, one would certainly observe an *average* increase in performance; however, since every subject will change an equal amount in the performance measure, one would *not* observe any *individual differences* in their change.

In this chapter and in the following chapter, we will be focusing primary attention on situations in which individual differences in changes are of interest. In Chapter 4, we will examine briefly a method for detecting whether average changes or trends are observable within subsets of our sample. Let us next turn our attention to three questions which guided the evolution of the proposed strategy.

What Part Should "Change Scores" Play in the General-Purpose Strategy?

An examination of this issue immediately immerses the investigator in some ticklish philosophical questions such as "What do we mean by change?," "How do we measure change?," and "Must we infer change rather than measure it?"[4] From these philosophical issues come procedural-statistical considerations involving the use of "raw difference" scores vs. various kinds of "adjusted gain" scores as measures of change. Chapters 2 and 3 will present a summary of the author's attempts to wrestle with these questions

---

[4]See especially Harris, 1967; Coleman, 1964b; Coleman, 1968; Cronbach & Furby, 1970.

and his recommendations vis-a-vis the use of change scores in longitudinal analyses of the Youth in Transition data.

A slightly different, but related, set of questions arises concerning the various ways in which change scores are put to use. For example, measures of change may be used to arrange individuals along a continuum for subsequent analyses. This implies that one or more new dynamic dependent variables (i.e., change scores) are being derived from a combination of static (i.e., cross-sectional) scores (as in Figure 1-2). Alternatively, a single, overall score might be calculated by somehow combining two or more of these change scores to summarize the amount and direction of an individual's overall shift during the interval of interest for each criterion dimension. This use of a change score might best be viewed as a variable reduction procedure because it results in a single dependent measure for each dimension rather than the ten measures previously discussed. Still another potential use for change scores is to aid in the identification of criteria where a good deal of change is taking place.

These three potential uses of change scores lead to somewhat different streams of analyses. However, they are interrelated to a substantial degree. For instance, if a method is found of deriving a change score which may be used as a dependent variable for subsequent analyses, then perhaps a single "overall change" score may be found for each criterion. These overall change scores might be used, in turn, to point to focal areas for early analyses by showing "where the action is." However, if a method for calculating change scores is not found, then questions regarding the other two potential uses of change scores may not be as readily addressed. It thus follows that the question of what kind of change score should be used is a crucial one to be considered early in the development of a general-purpose analytic strategy.

What Alternatives Exist to "Overall Change" Scores?

Since our attempt to develop a general-purpose strategy for longitudinal analyses leads us first to an examination of the role played by change scores, it is essential that alternatives to the use of such scores be carefully considered. The most promising alternative of this sort is what might be called "parallel prediction" of the static criterion scores. More specifically, this procedure calls for predicting from a selected set of important individual and environmental characteristics to the four static criterion measures. By noting whether or not a criterion is becoming "more predictable" across time, one may be able to infer that overall changes have taken place. And by noting which of the predictors are assuming greater explanatory power, one starts to get an indication of what kinds of variables may be influencing the inferred changes.

The second monograph in the Youth in Transition series (Bachman, 1970) may be viewed as a partial application of such a strategy. In summarizing the impact of the family background characteristics upon the initially measured criterion dimensions, a strategy was developed which first selected a limited number of important predictors from the much larger available set, and then used this limited set to predict separately to each of a set of selected criteria. Suppose we now performed parallel predictions from this same predictor set to the criterion dimensions measured at Times two, three, and four. If we observed in these analyses that the predictability of a criterion remained fairly constant over time, and if the relative importance of the predictors remained essentially the same, we would be inclined to conclude that this set of predictors has already established its pattern of influence on the criterion in question by the time of the initial measurement, and that this pattern is not changing during the period of observation. This conclusion does not follow *necessarily*; it is simply the most obvious and parsimonious conclusion. On the other hand, if a criterion were to become more predictable over time, there would be ample reason to consider concentrating additional efforts on discovering the ways in which certain predictors increased in their explanatory power.

By use of this kind of parallel prediction to the several static criterion scores (Times one, two, three, and four), it thus seems possible to identify criterion measures where change is taking place and at the same time to note which variables might be accounting for this change.

How Can Data from More Than Two "Waves" Be Used Most Efficiently?

The last of the three guiding questions used in the attempts to develop a general purpose strategy arose as the author was reviewing previous work in this area. With very few exceptions,[5] the overwhelming majority of previous efforts have been focused exclusively on problems of analyzing data from two-wave studies. It is not very surprising, therefore, that most of these previous efforts have limited utility for our four wave panel study. It seems intuitively obvious to the author that a general purpose strategy should make use of all available data.

At the practical level, we have six different intervals within which change could be examined. Should we focus attention on just one of these intervals? If so, which interval should be chosen?

[5]The most noteworthy exception is briefly summarized in Appendix A (the previously referred to Coleman chapter in Blalock and Blalock); but as we shall see later, this noteworthy exception contains other limiting features insofar as use with our data is concerned.

If not, how can we develop a general-purpose strategy that will apply to the 45 criteria which have been measured at four points in time?

To answer this question it will be helpful to present briefly one of the conclusions drawn from the analyses performed on the first three waves of data. While grappling with the question of what (if any) kind of change scores would best serve our purpose, it became obvious that the "parallel prediction" strategy outlined above worked well for pointing to where change was indeed taking place. Specifically, when a criterion dimension became more (or less) predictable across the Time 1 to 3 interval, then we selected that criterion for subsequent analyses aimed at evaluating various kinds of change scores. It is important to realize that this procedure permits the selection of criterion dimensions without involving many of the rather messy methodological complications which enter the picture whenever change scores of any type are used.

The fact that this procedure involves predicting to the repeated static scores indicates an additional advantage; these predictive relationships are often of considerable interest in their own right, whether or not change scores are utilized. The predictions to the initial criterion scores from the boys' family background characteristics and intelligence (Bachman, 1970) summarize the important family environment effects which were observed as the study began. Similarly, predicting to the Time 3 criteria from school characteristics (after appropriate "control" for individual background factors) will permit summaries of the effects of school environments.

Early application of the parallel prediction model thus seemed to make a good deal of sense. It serves well the purpose of pointing to areas where change may be taking place, and additionally produces analyses of considerable interest in their own right. Finally, and equally important, it makes use of the criterion data from as many waves as are available for analysis.

Summary

The number of both predictor and criterion measures available for analyses is so large that a general-purpose strategy is a necessity. In developing such a strategy for longitudinal analyses, an early and pivotal decision regards the use of change scores. Three potential approaches have been outlined, and each will be considered in analyses reported in subsequent chapters. The utility of various kinds of change scores will be evaluated in the context of an analytic strategy calling for "parallel prediction" to the repeated static scores. Such a strategy seems to be advantageous both because it points to areas where change has taken place and also because it makes efficient use of all criterion data available.

Chapter 2

CHANGE SCORES: SOME STRENGTHS AND WEAKNESSES

Before turning to an evaluation of the utility of change scores in longitudinal analyses of the Youth in Transition data in Chapter 3, it will be helpful to examine the solutions proposed by previous investigators in the area of measurement and analysis of change.[1] Although no single procedure has emerged from these previous efforts, there are some areas in which the authors seem to be essentially in agreement. Let us begin by examining these areas.

Areas of Agreement

Virtually all of the previous investigators agree that raw change or raw gain scores[2] are of questionable utility and can easily lead to fallacious conclusions. One reason for this limited utility derives from the commonly-observed negative

[1]See especially Harris, 1967; Coleman, 1968; and Cronbach & Furby, 1970.

[2]"Raw change" or "raw gain" is used to denote a derived score formed by simple subtraction of an earlier static score from the same measure obtained at a later point in time.

correlation between the initial static score and the raw gain score. (Parenthetically, we might note that in a parallel fashion, one would also observe a positive correlation between the final static score and the raw gain.) In the usual case in which the variances of the initial and final static scores are approximately equal, this negative correlation will be observed regardless of the sign of the relationship between the initial and final scores. (See Appendix B for a proof of this statement.) As a consequence of this negative relationship, other variables which are positively related to the initial score more than to the final score are also likely to show negative relationships with the raw gains. However, it is by no means clear that these other variables affected a "real" loss (negative change) in the criterion across the observed interval.

One of the problems with raw gain scores stems from the fact that they are systematically related to whatever amount of random measurement error is contained in the static scores. There again seems to be general agreement that issues of measurement error, although potentially important in all studies, are of critical importance in this area of measurement and analysis of change. Perhaps an example[3] will help to illustrate the scores of this problem. In Table 2-1, X represents

[3]This example is based on a discussion by Bereiter, 1967, p. 10.

the initial static score and Y represents the final score of the same measure. Next let r_X and r_Y be used to designate the internal consistencies (reliabilities) of the initial and final scores respectively, and let r_{XY} represent the stability across the interval calculated via the product-moment correlation coefficient between the initial and final scores. Of interest to us here is how various combinations of internal consistency and stability affect the internal consistency (reliability) of the derived raw gain score, represented by r_{Y-X}. The figures are based on the generally-observed fact that both the internal consistency and the variance of the static scores are constant across time.

TABLE 2-1

RELIABILITY OF RAW GAIN SCORES: AN ILLUSTRATION

Case	Internal Consistency	Stability	Internal Consistency of Raw Gain Score
I	$r_X = r_Y = .8$	$r_{XY} = .7$	$r_{Y-X} = .33$
II	$r_X = r_Y = .9$	$r_{XY} = .7$	$r_{Y-X} = .53$
III	$r_X = r_Y = .8$	$r_{XY} = .0$	$r_{Y-X} = .80$

Inspection of this table reveals a seemingly incomprehensible situation; namely, if one is interested in gain scores which have high reliability, he should apparently seek measures which have *high* internal consistency but *low* stability across time! But if he used such measures, how would he know whether or not he has measured the same thing at the two points in time? Bereiter's answer to this apparent paradox is one to which other authors seem to agree:

> Where it becomes crucial to decide whether or not one is measuring change is in the selection or construction of the measuring instruments. If one is measuring change, then it is as measures of change and only as measures of change that the validity and reliability of his instruments have any importance (Bereiter, 1967, p. 14).

Thus, the effect of measurement error in static scores from which raw change scores are derived is to decrease the reliability of the resulting change score. One should seek measures which are as reliable (i.e., internally consistent) as possible but which are not so stable across time that no change may be observed. This very important distinction between "split-half" vs. "test-retest" reliability has not received as much attention as it deserves. (See Heise, 1969, for an example of a thorough treatment of this distinction.)

In the next chapter, we will examine in considerable detail the internal consistency and stability coefficients for a set of criterion dimensions from the Youth in Transition study. Since these dimensions were selected before the above-mentioned suggestions were available, it will be of considerable interest to note whether the observed reliabilities and stabilities possess the previously-described characteristics to a degree sufficient to warrant using raw change scores for analyzing at least some of the dimensions. In this light it is of interest to note that Shaycoft concluded that raw gain scores based on Project TALENT's measures of aptitude and ability had reliabilities so low as to render them analytically useless (Shaycoft, 1967, pp. 4-19 through 4-30).

How to Improve on Raw Gain Scores

So far in this chapter we have noted areas of apparent consensus regarding some limitations in the use of raw gain scores. Unfortunately, there are almost as many solutions to the problems posed by these limitations as there are authors who have investigated the issues. A significant portion of the lack of agreement among the proposed solutions arises because of the many and varied potential purposes for gain scores. A partial list of such purposes would include the following:

- to provide a dependent variable for subsequent analyses
- to select exceptional individuals for additional study

- to obtain an indicator for a concept or construct in such a way that its relationship with other variables conforms most closely to a given theory
- to estimate change for an individual with respect to a group
- to examine mean changes for groups

Although the ingredients of this list are not intended to be mutually exclusive, they perhaps serve to illustrate the importance of keeping in mind the purpose(s) which gain scores will serve once they have been derived.[4]

On the Youth in Transition study, the primary interest in deriving change scores relates to their use as dependent variables in subsequent analyses. Therefore, let us examine a few of the previously suggested procedures for improving on raw gain scores as dependent variables.

The Lord Procedure (see Lord, 1967, pp. 21-38). The Lord procedure considers an "initial" score X and a "final" score Y applied to each of a sample of subjects on two occasions. True scores X_T and Y_T for each individual at these times are postulated, following directly from similar formulation in classical test theory. The procedure is aimed at estimating a true difference or gain

[4]The reader may recall from the first chapter that at least three different, but related, potential uses of change scores are being considered in the development of this general-purpose strategy. (See pp. 7-8.)

score $G_T = Y_T - X_T$ for each individual via the determination of regression coefficients for an equation of the form: $\hat{G}_T = \overline{G} + \beta_{GX \cdot Y}(X-\overline{X}) + \beta_{GY \cdot X}(Y-\overline{Y})$.[5] (1) Lord shows that, for a sufficiently large number of cases, the average true gain, $\overline{G}$, may be estimated by the arithmetic difference between the *observed* means; namely $\overline{G} = \overline{Y} - \overline{X}$. The estimated true gain (Formula 1) then follows from assuming that the variances of the initial and final scores are equal (as is typically the case) and that the errors are uncorrelated with the true scores. The procedure then reduces to one of estimating the two partial regression coefficients as follows:

$$\beta_{GX \cdot Y} = \frac{(1-r_Y)\dfrac{r_{XY}s_Y}{s_X} - r_X + r_{XY}^2}{1-r_{XY}^2}, \text{ and} \qquad (2)$$

$$\beta_{GY \cdot X} = \frac{r_Y - r_{XY}^2 - (1-r_X)\dfrac{r_{XY}s_X}{s_Y}}{1-r_{XY}^2}. \qquad (3)$$

As before, r_X and r_Y represent the reliabilities of the initial and final scores respectively. These reliability coefficients are first estimated

[5]In these and following discussions, a bar above a term designates the average value across individuals and β will be used to represent standard partial regression coefficients via the usual subscript notation showing the correlated residualized variables before the dot and the variable used in obtaining the residuals after the dot.

by an independent procedure and then, along with the observed stabilities and standard deviations, are substituted in equations 2 and 3 to solve for the regression coefficients. These coefficients are in turn substituted into equation 1 (along with the difference between the observed means) in order to solve for an individual's true gain score.

In summary, Lord's procedure permits the estimation of an individual's true gain, using the individual's observed initial and final scores and statistics based on the total set of observed individuals.[6]

The Bereiter Procedure (see Bereiter, pp. 3-20). Whereas the major objective of the Lord procedure was to derive an individual measure of true gain so that this measure can then serve as a new dependent variable to be predicted from other characteristics of interest, the Bereiter procedure is aimed at estimating such relationships (i.e., between predictors and true gain scores) directly. Specifically, to obtain the correlation between an independent (predictor) variable W and the final static score Y in a way that "controls for" the individual's score on the initial dependent variable X, he suggests using the following formula:

[6]For a critical review of the underlying assumptions, see Cronbach & Furby, 1970, pp. 69-70.

$$r_{WY \cdot X_T} = \frac{r_{WY} - \dfrac{r_{WX} r_{XY}}{r_X}}{\sqrt{r_X - r_{WX}^2}\,\sqrt{r_X - r_{XY}^2}}. \tag{4}$$

Note that it is X_T, the initial *true* score, which is being partialled out of the WY relationship in the formula. This indicates the importance to Bereiter of correcting for the unreliability in the initial score. Because the sign of the denominator in Formula 4 will always be positive, it is the sign of the numerator which determines whether the estimated relationship will be positive or negative.

Bereiter shows that if the initial *raw* score (rather than the true score as in Formula 4) is partialled out, the numerator of the estimated relationship is $r_{WY} - r_{WX} r_{XY}$. Thus, taking into account the initial score reliability could actually reverse the sign of the relationship based on raw score calculations. Therefore, whether or not one corrects for this unreliability has potentially important implications. This decision is not an empirical one; rather, it follows from the analyst's decision as to whether he seeks a set of change scores orthogonal to the initial *observed* scores or to the estimated initial *true* scores.

Bereiter next addresses the question as to whether or not a similar correction for unreliability in the final Y scores should be made. It would seem consistent with our objective (i.e., to estimate the relationship between the independent variable W and $Y_T - X_T$, the true gain) to make such a correction. Bereiter's argument is that this adjustment is not critical. This follows from the fact that the adjustment takes the form of replacing r_X with the product $r_X r_Y$ in the second factor of the denominator of Formula 4. Whenever r_Y is less than one (and it can never exceed 1) the effect of this change is to produce a smaller denominator which, in turn, results in a larger absolute value for the estimated relationship. Unlike the correction for initial score unreliability, however, the direction of the two estimated relationships ($r_{WY_T \cdot X_T}$ and $r_{WY \cdot X_T}$) will always be the same; one will simply observe that $r_{WY_T \cdot X_T}$ will usually be larger than will $r_{WY \cdot X_T}$. From this Bereiter concludes that the choice between these two coefficients is not critical.

In short, then, the Bereiter procedure may be used to estimate directly the relationship between a predictor variable, W, and true gain score, $Y_T - X_T$, without ever estimating the true gain score itself. As in the Lord formulae, initial and final observed scores at both individual and average levels, as well as observed interrelationships among these scores, are used in obtaining the estimated relationship. His derivations suggest that the practical effect of adjusting these estimates for final score unreliability is relatively unimportant compared to the adjustment for initial score unreliability.

The Cronbach-Furby Procedure (see Cronbach & Furby, 1970). Like the Bereiter procedure, Cronbach and Furby suggest a means of estimating relationships between predictor variables and estimated true gains without actually calculating the true gain scores. In fact, they conclude that

> ...gain scores are rarely useful, no matter how they may be adjusted or refined (Cronbach & Furby, 1970, p. 68).

In spite of this stand, they propose a procedure for estimating true gains, both because they feel it provides a better estimator to use in the limited cases where they recommend using change scores, and because

> Very likely some investigators will decide to obtain change or difference scores, even for problems where we consider such measures inappropriate. Such a person will often find one of our estimation formulas better than those now suggested in the literature (Cronbach & Furby, 1970, p. 68).

The Cronbach-Furby discussion presents an important extension of the Bereiter model in that (a) several W (predictor) variables may be examined simultaneously (as would be the case, for example, in multiple linear regression models), (b) their proposed model is appropriate for analyses of differences between two variables in a cross-sectional study as well as for longitudinal analyses, and (c) it introduces a new class of Z variables; namely, those variables measured at the time of (or after) the final measure, but which might be used to further refine the estimate of true gain.

With respect to the Youth in Transition project, extensions b and c above offer little if any help. However, the ability to handle several predictor variables at once represents a potentially important addition to previously available procedures.

In addition, the authors demonstrate that if one's objective is to identify individuals who have gained (or lost) an exceptional amount, then the individual's true residual gain need not be calculated. Rather, the "raw residual-gain score," $D{\cdot}X = Y-\bar{Y}-\beta_{Y.X}(X-\bar{X})$,[7] is well suited for such a purpose. (Notice that this score is *not* equivalent to the raw gain score.) However, if the individual's true residual gain is to be estimated, the authors provide a different method for estimating it.

Thus, the Cronbach-Furby paper presents an extension of the previously discussed Bereiter procedure via a procedure to be used for estimating relationships between true gain scores and a set of dependent variables. In addition, the authors suggest the use of raw residual-gain scores

[7]Formula 21 (Cronbach & Furby, 1970, p. 74) appears to be correct in this regard. A "correction" (Errata, 1970, p. 218) is in error.

when the objective is to identify exceptional gainers or losers within a group of individuals.

Summary

All three of the procedures discussed above (Lord, Bereiter, and Cronbach-Furby) deal with methods of improving on raw gain scores as dependent variables. This improvement was seen to be necessary because of the regularly observed negative correlation between raw gains and initial scores, the positive correlation between raw gains and final state scores, and also because of the confounding of random measurement error with the raw gain scores. The three procedures outlined do not all address themselves to the same objectives. This illustrates the need for the analyst to identify carefully the purpose(s) of the gain score he seeks before choosing a procedure for producing such a score.

It is worth noting again here that none of the procedures discussed make efficient use of data from more than two points in time; thus, they all fail to meet one of the objectives put forth in the first chapter as a desired condition for our general-purpose strategy. Coupled with the desirability of predicting the static criterion scores in order to achieve analytic objectives of considerable importance in their own right,[8] the failure of gain scores to meet this

[8]See Bachman, 1970, for an example of a report based on analyses of this sort.

objective casts even further doubt upon their general utility in our longitudinal analyses model.

Chapter 3

STABILITY VS. CHANGE IN THE YOUTH IN TRANSITION DATA

This chapter is devoted to an evaluation of change scores for longitudinal analyses of the Youth in Transition data. As an integral part of this evaluation, adjusted gain scores are compared with raw change scores, with an eye to noting whether the advantages of adjusted gains described in the previous chapter actually are observable in the available data.

Stability in the Criteria

Table 3-1 presents the means and standard deviations of 18 criterion dimensions[1] measured at each of the first three data collections. (For reasons described in Chapter 1, these analyses are based on the 1,374 boys who stayed in the same school throughout their sophomore to senior years.)

The data in Table 3-1 allow us to investigate whether or not the school environment is bringing about consistent changes in our criterion dimensions. If the school were exerting an important

[1]See Bachman, et al., 1967, Chapter 4 for a description of the composition of these measures.

TABLE 3-1

TIME 1, 2, AND 3
MEANS AND STANDARD DEVIATIONS
OF 18 SELECTED CRITERIA

	Time 1		Time 2		Time 3	
	$\overline{X}$	S.D.	$\overline{X}$	S.D.	$\overline{X}$	S.D.
Job Information Test	17.11	3.22	18.14	3.34	18.96	3.25
Positive School Attitudes	3.30	.49	3.21	.51	3.06	.54
Negative School Attitudes	1.80	.55	1.79	.53	1.90	.55
Need for Self-Development	3.66	.51	3.64	.49	3.65	.48
Need for Self-Utilization	3.89	.51	3.82	.49	3.85	.48
Self-Esteem	3.77	.51	3.84	.48	3.88	.49
Negative Affective States	2.59	.54	2.54	.53	2.53	.54
Happiness	3.80	.60	3.82	.60	3.80	.59
Somatic Symptoms	2.08	.54	2.07	.53	2.09	.52
Social Values	4.75	.54	4.79	.48	4.76	.45
Ambitious Job Attitudes	5.15	.65	5.30	.63	5.31	.63
Internal Control	1.67	.19	1.71	.20	1.71	.21
Trust in People	1.52	.37	1.54	.39	1.52	.40
Trust in the Government	3.70	.65	3.54	.61	3.51	.59
Delinquent Behaviors	1.54	.42	1.51	.43	1.57	.41
Academic Achievement (Grades)	41.23	6.83	40.67	6.83	41.25	7.11
College Plans	.64	.48	.67	.47	.57	.50
Occupational Aspirations	63.78	25.47	61.09	24.17	59.77	24.15

and consistent influence on these criteria, we might expect to find the means in Table 3-1 moving in the same direction across time. If the school exerted a facilitating effect, the Time 3 mean would be expected to be larger than the Time 2 mean which, in turn, should be larger than the Time 1 mean. If the school exerted a debilitating effect, just the opposite pattern should be observed; that is, we would expect the means to drop across time.

A second type of school effect may be observed by examining the standard deviations in Table 3-1. If schools were causing students to become more alike, then the scores would tend to converge more as time passed. This convergence would be indicated by a "shrinkage" in the standard deviations at subsequent data collections. On the other hand, if schools were encouraging students to become less alike (as might be the case in schools which truly developed the student's individuality, for example), then we would expect to find that the standard deviations are increasing across time. This second type of school effect bears directly on whether schools are acting as a "conforming" or a "non-conforming" agent.

When the data in Table 3-1 are examined, the overall picture which emerges is one of considerable stability, both in the means and in the standard deviations. We shall see later in this chapter that this stability exerts an important effect on the potential use of change scores.

A few of the dimensions in Table 3-1 show evidence of systematic shifts in the means.[2] However, none of these shifts appears to be very large. As a matter of fact, Job Information is the only dimension to evidence a shift in which the mean gain equals or exceeds one-quarter of a standard deviation across both intervals. And none of the dimensions demonstrate large convergence or divergence in their scores. At this point, then, we have seen little evidence of either type of school effect in these criterion dimensions.

Another way to examine the stability of measures repeated at two or more points in time is through the use of correlation coefficients. Instead of asking whether or not there are shifts in the means and/or variances, this second kind of investigation focuses on whether individuals change across time relative to the other individuals in the sample. As more and more individuals hold the same relative position through time, the correlation between the scores at the beginning and end of an observed interval will approach unity.

In examining such "stability coefficients," we cannot ignore the potential effects of

[2]Job Information, Self-Esteem, Ambitious Job Attitudes, Internal Control, and Trust in the Government show increases across both intervals, and Positive School Attitudes shows a consistent decrease.

measurement error. Operationally, the issue may be thought of as trying to distinguish whether the lack of perfect stability is due to lack of perfect reliability in the measuring instrument or whether "real" shifts have taken place. Having data from three points in time helps to make this distinction. In a panel study of political attitudes during the 1956, '58, and '60 elections (Converse, 1963), Converse discusses some of the possible implications of the stability of his data from three waves.

> The most revealing statistical property of these attitude-change data emerges when we consider not simply the correlations between the same attitudes over two-year spans, but also the correlation for each attitude between the initial and terminal interviews, a span of four years. For we discover that these t_1-to-t_3 correlations tend to be just about the same magnitude as the t_1-to-t_2 correlations, or the t_2-to-t_3 correlations. That is, surprising though it may be, one could predict the 1960 attitudes on most of these issue items fully as well with a knowledge of individual attitudes in 1956 alone as one could with a knowledge of the more proximal 1958 responses. Furthermore, the tendency toward parity of the three correlations is clearest among the issue items with greatest turnover; among the more nearly stable items, the four-year correlation tends to be slightly lower than the two-year correlations, a pattern which is of course much closer to our intuitive expectations (Converse, 1963, pp. 7-8).

Let us consider a model in which an observed score is comprised of a true score plus a random measurement error component. Let us further assume that in the observed population the error components are distributed normally around a zero mean, are uncorrelated with the true score, and are uncorrelated across time. Now if the true scores were perfectly stable, the observed correlations would vary from unity because the error component varies. However, since the error components are serially uncorrelated, the observed score correlations would not be expected to vary with the length of the interval. Such a model could account nicely for the Converse data.

On the other hand, if real change in relative position were taking place (i.e., if the true scores were not perfectly stable), then we would expect to observe lower stabilities for longer intervals and higher stabilities for shorter intervals. This is due to the fact that when real change is occurring, the longer the interval, the more reordering or changing we would expect.

Table 3-2 presents the stability coefficients for the 18 dimensions contained in Table 3-1. Again, the overall picture is one of considerable stability. However, the data in this table are consistent with the idea that some real changes in our criteria may be taking place during the high school years. This follows from the observation that the highest stability coefficients are found in the column corresponding to the shortest interval (Time 2 to 3), the lowest coefficients are located in the column corresponding to the

TABLE 3-2

TIME 1-3, 1-2, AND 2-3 CORRELATIONS
FOR 18 SELECTED CRITERIA

	Time 1-3 (30 mos)	Time 1-2 (18 mos)	Time 2-3 (12 mos)
Job Information Test	.53	.59	.61
Positive School Attitudes	.42	.49	.57
Negative School Attitudes	.41	.47	.54
Need for Self-Development	.50	.56	.64
Need for Self-Utilization	.42	.48	.54
Self-Esteem	.49	.54	.66
Negative Affective States	.52	.56	.70
Happiness	.47	.54	.63
Somatic Symptoms	.42	.52	.62
Social Values	.41	.51	.54
Ambitious Job Attitudes	.36	.46	.52
Internal Control	.32	.42	.51
Trust in People	.35	.37	.47
Trust in the Government	.33	.46	.48
Delinquent Behaviors	.48	.53	.63
Academic Achievement (Grades)	.58	.67	.66
College Plans	.40	.44	.49
Occupational Aspirations	.53	.62	.66

longest interval (Time 1 to 3), with the third set of stabilities (corresponding to the Time 1 to 2 interval) falling between these two extremes.

In the second chapter we noted that both the stability and the internal consistency of a measure exert important influences on the reliability of the raw change score. We have just seen that our criterion dimensions are relatively stable during the periods where we have observed them. Let us now proceed in our evaluation of change scores by examining the internal consistencies of a selected set of criteria.

Internal Consistencies of the Static Scores

The internal consistency coefficients are important indicators of the quality of the criterion data. In addition to yielding information about the reliabilities of our measures, such coefficients may greatly aid our attempts to detect and to understand changes. In Table 2-1 we saw that if a change score is to be optimally useful, it should be based on static scores whose reliability is as high as possible. Ideally, we seek measures whose reliability remains approximately constant across time so that the same amount of measurement error exists in the final scores as existed in the initial scores, thus increasing the potential utility of change scores derived from these static scores.

The internal consistency estimates reported here are Cronbach-Alpha coefficients (Cronbach, 1951, pp. 297-354). This reliability estimate was chosen because it is well suited to our rating-scale type of data. Additionally, it may be interpreted directly as the proportion of true score variance accounted for by the observed score (Nunnally, 1967, p. 196), a characteristic which makes it especially useful for our present purposes. Finally, but also very important, an efficient procedure for estimating the coefficient from item responses is available within our own software system.[3]

Table 3-3 presents the reliability estimates for the six dimensions identified in Tables 3-1 and 3-2 as potentially reflecting the greatest systematic shifts from waves 1 to 3. It is immediately apparent that, except for Internal Control, no large shifts are found in the internal consistencies from Time 1 to Time 3. Equally observable is the fact that these measures have reasonably good reliabilities; an average of 71 percent of the true score variance being accounted for by the observed scores for the six dimensions in this table.[4]

---

[3] This program is based on a paper by Bohrnstedt, 1969, pp. 542-548.

[4] Because of the lack of "balance" (i.e., all 15 items are reversed) in the Positive School Attitudes scale, response bias is very likely leading to an overestimate of the internal consistencies.

TABLE 3-3

RELIABILITY ESTIMATES (CRONBACH α)
FOR SIX SELECTED CRITERIA

Dimension	# Items	# Reversals	# Choices in Response Scale	T1	T3
Job Information Test	25	--	2-5*	.699**	.671**
Self-Esteem	10	6	5	.737	.785
Positive School Attitudes	15	15	4	.909	.912
Ambitious Job Attitudes	13	6	4	.637	.640
Internal Control	12	5	2	.554	.675
Trust in the Govt.	3	1	5	.637	.635

*Recoded to "right-wrong" versions for estimate

**Item responses for this test were scored for correctness, and the recoded answers were then input to the Kuder-Richardson Formula 20 reliability estimate. (See Nunnally, 1967, pp. 196-197.)

With respect to the criterion dimensions examined to this point, we find evidence of a good deal of stability and of internal consistency. Let us now turn our attention to the implications of these empirical observations for the use of change scores.

Effects of Static Score Stability and Internal Consistency on Change Scores

Earlier we saw that when the stability of a static score was fairly high, the internal consistency of the raw change score derived from the static scores would be low. Then we saw that most of our criterion dimensions had relatively high stability coefficients. The primary purpose of this section is to compare the reliabilities of adjusted gain scores to those of raw change scores. There is a very pragmatic objective which urges that this comparison be made; namely, if the theoretical advantages which previous authors suggest should accompany the adjusted gain scores are not observable in our data, then we see little reason to go to the considerable expense of performing the adjustment. If the advantages of adjusting are obtained, however, then we will want to proceed into the next stage of analysis with such adjusted gain scores, and not with raw difference scores.

Following are formulae for the reliability coefficients of raw difference, independent gain, and regressed gain scores.[5] Here X and Y will be used to represent the initial score and the final score respectively, r_X and r_Y the internal consistency coefficients, and r_{XY} the stability coefficient across the interval of observation.

$$\text{Reliability of Raw Difference} = \frac{r_Y s_Y^2 - 2r_{XY} s_X s_Y + r_X s_X^2}{s_Y^2 - 2r_{XY} s_X s_Y + s_X^2} \qquad (5)$$

$$\text{Reliability of Independent Gain} = \frac{r_X (r_X r_Y - r_{XY}^2)}{r_X^2 - 2r_{XY}^2 r_X + r_{XY}^2} \qquad (6)$$

$$\text{Reliability of Regressed Gain} = \frac{r_Y - 2r_{XY}^2 + r_{XY}^2 r_X}{1 - r_{XY}^2} \qquad (7)$$

Applying these three formulae for reliability to the overall (Time 1 to 3) change scores for the six dimensions previously discussed yields estimates displayed in Table 3-4. The internal consistencies, stabilities, and standard deviations of the static initial and final scores are represented along with the three types of change

[5]These formulae are taken from Tucker, et al., 1966, pp. 468-469, but the notation has been modified so as to be consistent with that used in Chapter 2.

TABLE 3-4

TIME 1 TO 3 RELIABILITIES FOR SIX SELECTED CRITERIA

Dimension	Internal Consistencies r_x	r_y	Stability r_{xy}	Deviations s_x	s_y	Reliabilities of Change Scores: Observed Differences	Independent Gains	Regressed Gain
Self Esteem	.737	.785	.49	.51	.49	.53	.58	.63
Job Information Test	.699	.671	.53	3.22	3.25	.33	.35	.42
Positive School Attitudes	.909	.912	.42	.49	.54	.85	.87	.87
Ambitious Job Attitudes	.637	.640	.36	.65	.63	.44	.48	.53
Internal Control	.554	.675	.32	.19	.21	.44	.51	.58
Trust in the Government	.637	.635	.33	.65	.59	.46	.50	.55

score reliability estimates. It is apparent that there is very little to be gained from adjusted or regressed gain scores, at least insofar as reliability increases are concerned. In addition, even reliabilities of the regressed gain scores are not very high, averaging .60 for these six dimensions. Excluding the Positive School Attitudes scale (because of the confounding of response bias with internal consistency), this average reliability drops to .55. When we recall that these dimensions were selected because of their likelihood of showing change, we get an even more pessimistic picture of the utility of these change scores.

The Predictability of Change Scores

At this point in the development of the analytic strategy, the use of change scores was becoming increasingly doubtful. Before deciding to abandon them altogether, however, a series of analyses was conducted to test their predictability. Previous analyses had already demonstrated that many of these criterion dimensions were equally predictable (at Times 1, 2, and 3) from background characteristics (Bachman, 1970, p. 208ff.). Thus, it seemed highly unlikely that change scores would be predictable in these instances where the background effects are so stable during the high school years.

The boys' occupational aspirations were singled out for these analyses both because there was a consistent drop in the average measure across time and because of the fact that this measure is very likely to be highly reliable.[6] In addition to these factors, it is an extremely important criterion dimension and one which is likely to be affected by things happening in the high school environment.

The sequence of the analyses was to first predict in a bivariate fashion to the Time 1, 2, and 3 static scores from a set of important background characteristics. Then, raw difference and independent gain scores were predicted from the same set of characteristics. Finally, joint predictions were made from the set of background characteristics to the static, raw difference, and independent gain scores.[7]

The results of these analyses are displayed in Table 3-5. The stability of the pairwise and joint relationships to the static scores may be observed by reading the coefficients in the first three columns. The next three columns attest to

[6]The method of producing the status of aspired occupation score is described in Bachman, 1970, pp. 173-174. Because it is a single question, no internal consistency coefficient could be calculated for this dimension.

[7]These joint predictions utilized a multiple regression technique called Multiple Classification Analysis (MCA), developed by Andrews, et al., 1967. See Bachman, 1970, pp. 62-75 for an application of this technique using Youth in Transition data.

TABLE 3-5

BIVARIATE[1] AND MULTIVARIATE[2] PREDICTION OF STATUS OF ASPIRED OCCUPATION; STATIC, RAW CHANGE, AND INDEPENDENT GAIN SCORES

	Static Score			Raw Change Score			Independent Gain Score		
	Time 1	Time 2	Time 3	3-1	3-2	2-1	3-.5(1)	3-.8(2)	2-.5(1)
a. Quick Test	33 11	31 10	33 11	02 --	05 --	05 --	21 04	12 01	20 04
b. Race	10 01	09 01	09 01	05 --	07 01	07 01	09 01	09 01	07 01
c. Number of Siblings	24 06	20 04	22 05	09 01	09 01	10 01	14 02	11 01	15 02
d. S.E.L.	35 12	29 09	32 10	05 --	08 01	08 01	16 03	12 01	16 03
e. Unadjusted R/R^2 from MCA	44 20	39 16	42 17	12 02	16 03	17 03	25 06	20 04	26 07
f. Adjusted R/R^2 from MCA	42 18	37 14	40 16	09 01	06 --	08 01	20 04	13 02	21 05

(If the table entry is less than 01, a dash will be used. For all other cell entries, the decimal point has been omitted for ease of reading; the numbers should be read as hundredths [i.e., as if there were a leading decimal point].)

[1]The first four rows of the table contain eta values in the upper left and eta-squared values in the lower right corners of each cell.

[2]The last two rows of the table contain multiple R values in the upper left and multiple R-squared values in the lower right corners of each cell.

the fact that the corresponding relationships to the raw change scores are extremely small, as we expected. Of particular note, however, are the magnitudes of the coefficients in the last three columns. Except for race, the eta coefficients all exceed .10, suggesting that relationships do indeed exist between the independent gain scores and the background factors.

These apparent relationships are most troublesome, because we had reason to believe from the coefficients in the first three columns that these background characteristics would not be related to changes in the aspired occupational status. The problem may be resolved when we examine more closely the nature of the independent gain score being analyzed here.

Independent gain scores of this type are aimed at adjusting the raw difference score for what is typically called a "regression effect." This effect is due to the commonly observed situation in which extreme scores at the initial observation "regress" toward the mean (i.e., they tend to be less extreme) at the time of the final observation. The raw difference score thus "penalizes" a respondent with a high initial score by increasing the probability that he will have a lower final score (and thus a negative raw change). The independent gain score adjusts for this effect by not subtracting all of the initial score out.

Specifically, these scores take the form

$$\hat{G} = Y - \beta_{Y \cdot X}X \qquad (8)$$

where X and Y are, as before, the initial and final scores, and $\beta_{Y \cdot X}$ is the regression coefficient of the final score on the initial score.[8] Since $\beta_{Y \cdot X}$ will, in the usual case, be a value less than unity, the resulting regressed gain score will be comprised of the final score minus *a part of* the initial score. In the limit, when $\beta_{Y \cdot X} = 1$, the independent gain scores and the raw difference scores will be identical. In the other limit (i.e., where the stability coefficient is zero), $\beta_{Y \cdot X} = 0$, and the independent gain score equals the final score.

In most situations, the regression coefficient will be neither 0 nor 1, but somewhere between.[9] Adjusting in this way thus produces a set of scores which are less negatively correlated with the initial score than are raw difference scores; but these adjusted gain scores will be more positively correlated with the final scores than are the raw gain scores. Thus, what we are observing in the right three columns of Table 3-5 is, at least in part, this undesirable feature of

---

[8]Algebraically, $\beta_{Y \cdot X} = r_{XY}\frac{(s_Y)}{(s_X)}$ (Hays, 1963, p. 504).

[9]Regression coefficients can be negative, but this is highly unlikely in the present instance. It would only obtain when a variable correlated negatively with itself across time. In Table 3-2, the lowest correlation observed was .32.

independent gain scores. To the extent that a predictor variable relates to a final score in about the same degree as to the initial score, it will also show an artifactual relationship to this kind of independent gain score.

This is one of major reasons that Lord did not define his regressed gain scores via equation eight. Rather, he defined a set of scores which would be unrelated to both the initial (true) and final (true) scores. He accomplished this by defining estimated true gain as (Lord, 1967, p. 28):

$$\hat{G} = \bar{G} + \beta_{GX\cdot Y}(X-\bar{X}) + \beta_{GY\cdot X}(Y-\bar{Y}). \qquad (9)$$

In this formula, the mean gain ($\bar{G}$) may be estimated (for a sufficiently large sample) by the mean difference ($\bar{Y}-\bar{X}$), and the regression coefficients (adjusted for measurement error) may be estimated as follows:

$$\beta_{GX\cdot Y} = \frac{(1-r_Y)\frac{r_{XY}s_Y}{s_X} - r_X + r_{XY}^2}{1 - r_{XY}^2}, \text{ and} \qquad (10)$$

$$\beta_{GY\cdot X} = -\frac{(1-r_X)\frac{r_{XY}s_X}{s_Y} - r_Y + r_{XY}^2}{1 - r_{XY}^2}. \qquad (11)$$

As before, r_X and r_Y represent the internal consistencies of the initial and final scores, s_X and s_Y the standard deviations of the initial and final scores, and r_{XY} the stability coefficient across the interval for which the adjusted gain score is being calculated.

This procedure outlined by Lord thus avoids the undesirable "built-in" relationship between the final score and a regressed gain score of the form $\hat{G} = Y - \beta_{Y \cdot X}X$. Let us examine the Lord procedure in more detail in order to determine whether such adjusted gain scores would be useful.

From Tables 3-1 and 3-3 we observed the following things about most of the criterion dimensions.

(A) The means for any one dimension were very stable across time. From this the following approximation holds: $\bar{X} = \bar{Y} \equiv M$. (12)
From this it follows that:
$\bar{G} = \bar{Y} - \bar{X} = 0$. (13)

(B) The standard deviations for a repeated measure did not change very much across time. Thus, it will generally be true that: $s_X = s_Y$, and from this it (14)
follows that: $\frac{s_X}{s_Y} = \frac{s_Y}{s_X} = 1$. (15)

(C) The internal consistencies for most of the criteria did not change from Time 1 to Time 3. Thus, $r_X = r_Y \equiv r$. (16)

Now, substituting appropriately from equations 15&16 into equations 10 and 11 yields the following:

$$\beta_{GX \cdot Y} = \frac{(1-r)r_{XY} - r + r_{XY}^2}{1 - r_{XY}^2}, \text{ and} \qquad (10.1)$$

$$\beta_{GY \cdot X} = - \frac{(1-r)r_{XY} - r+r^2_{XY}.}{1 - r^2_{XY}} \qquad (11.1)$$

From equations 10.1 and 11.1 it follows that

$$\beta_{GX \cdot Y} = - \beta_{GY \cdot X}. \qquad (17)$$

Substituting from equations 12, 13, and 17 into equation 9 for the estimated true gain, we see that:

$$\hat{G} = \beta_{GX \cdot Y}(X-M) + (-\beta_{GX \cdot Y})(Y-M).$$

Expanding, we get:

$$\hat{G} = \beta_{GX \cdot Y}X - \beta_{GX \cdot Y}M - \beta_{GX \cdot Y}Y + \beta_{GX \cdot Y}M.$$

Hence, $\hat{G} = \beta_{GX \cdot Y}(X-Y).$ (18)

In terms of the other regression coefficient we have an equivalent statement; namely,

$$\hat{G} = \beta_{GY \cdot X}(Y-X). \qquad (19)$$

Let us look at a rather typical example where the internal consistency ($r = .7$) and the stability ($r_{XY} = .5$) coefficients are about average for our data.

$$\beta_{GX \cdot Y} = \frac{(1-r)r_{XY} - r+r^2_{XY}}{1 - r^2_{XY}}$$

$$\beta_{GX \cdot Y} = \frac{(1-.7).5 - .7+.75}{1 - .25}$$

$$\beta_{GX \cdot Y} = -.47$$

$$\hat{G} = \beta_{GX \cdot Y}(X-Y)$$

$$\hat{G} = -.47(X-Y)$$

$$\hat{G} = .47(Y-X).$$

So, in this example, each individual's estimated true gain score would be calculated by multiplying his raw difference score by .47.

The relationships given in equations 12-19 will be good approximations for most of our criterion dimensions. It will in general be true that following Lord's procedure to estimate true gain will yield a score which is simply a constant multiplied by the raw difference score. Thus, their correlations with the raw difference scores will be very close to 1. And since the constant multiplier will always be less than one, the estimated true gain scores will have smaller variances, making it very unlikely they will be more predictable. From this information, we conclude that estimated true gain scores have very limited utility for analyzing the majority of the Youth in Transition criterion data.

Summary

In this chapter we have examined a large number of criterion dimensions which were measured at each of the first three points in time. In general, these dimensions were characterized by a great deal of stability in their means and variances. Also, their auto-correlations suggested that most of the sample were maintaining their relative position on these dimensions throughout the high school years. In addition, we examined internal consistency measures for a set of dimensions; they indicated that the measures were fairly reliable and that the reliability did not shift much across time.

The remainder of the chapter was devoted to an exploration of the effects of this general stability and reliability on the use of change scores in the analysis of these repeatedly measured criteria. We witnessed the fact that independent gain scores and regressed gain scores have higher reliabilities than do raw difference scores, but the increase in reliability is by no means large. Furthermore, we noticed that the independent gain scores were predictable in a situation where we expected no prediction to be observed. We found that this was due, at least in part, to an artifactual relationship between the independent gain and final (static) scores arising from the method used to calculate the independent gain score.

We next examined a procedure suggested by Lord (and outlined in an earlier chapter) to derive a true gain score which did not show this undesirable relationship to the final static score. When the conditions following from the earlier-noted stability and reliability of our criteria were reintroduced in conjunction with Lord's procedure, we observed that the resulting estimated true gain scores would not order our sample any differently than would raw difference scores, and that the estimated true gain scores would have smaller variances.

Therefore, we concluded that for the purpose of deriving a dependent variable for subsequent analyses, none of the adjusted gain score procedures result in sufficient improvement over the raw difference score to warrant the considerable effort and expense necessary to calculate them.

Chapter 4
EVIDENCE OF TRENDS AND SUBGROUP ANALYSES

To this point, we have seen little evidence that change scores, raw or residualized, will facilitate the longitudinal analyses of the Youth in Transition data. It is the case, however, that changes in many of our criteria are occurring in at least some of the boys in our sample.[1] This raises the two additional questions to be addressed in this chapter. Is there any evidence of trends in the scores of those boys who do indeed change? How shall we identify and analyze those sets of respondents that show particular patterns of change across time? Let us now turn our attention to the first of these questions.

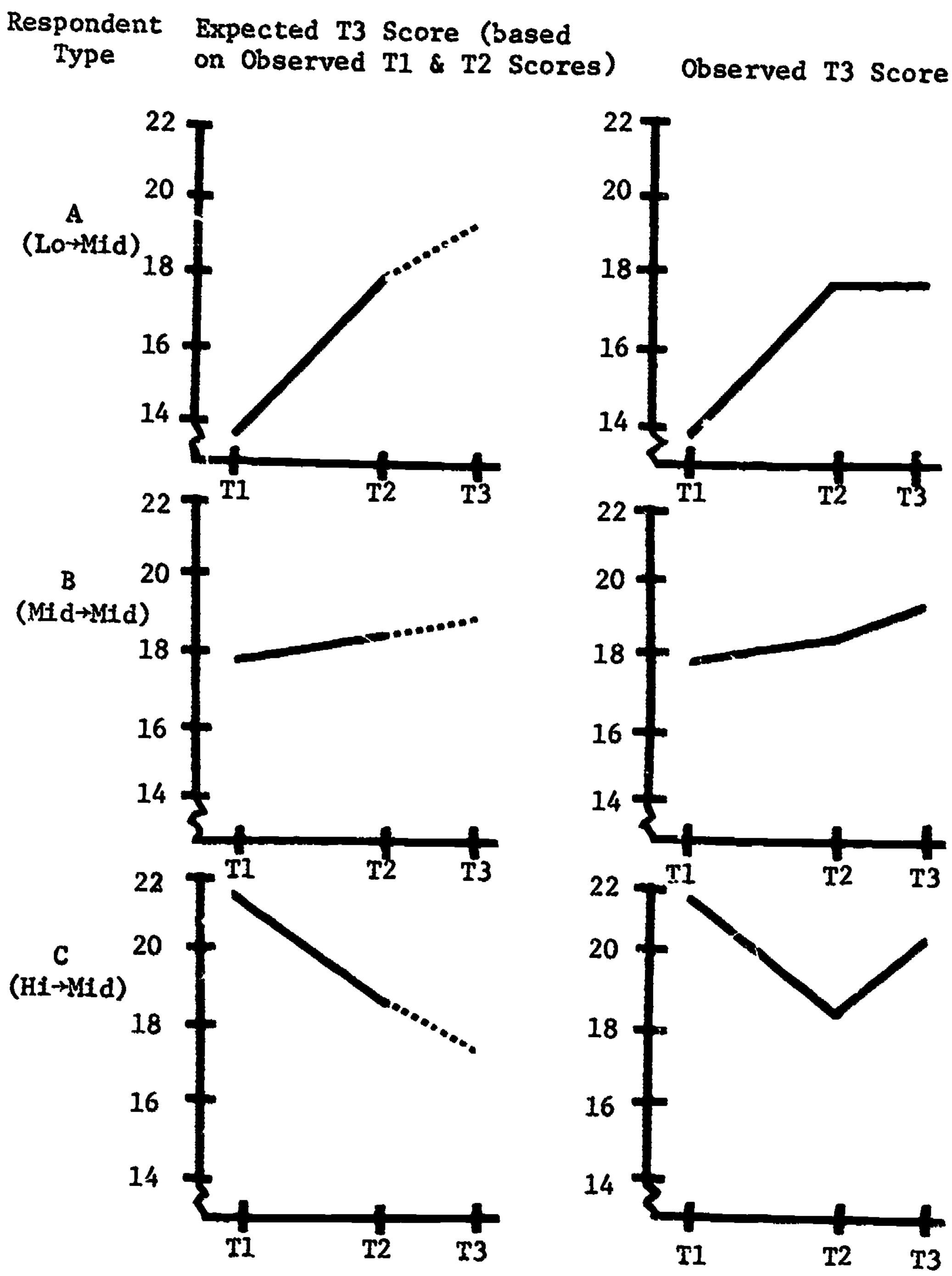
Evidence of "Trends" in the Observed Changes

Before discussing the statistical procedure for indicating trends in the data, let us first examine what is meant by a trend. Consider three respondents all of whose T2 scores are in the middle bracket of the trichotomy. (See Figure 4-1.)

[1]For example, the stability coefficients in Table 3-2, while quite high, still leave room for quite a bit of change to be taking place across each interval.

FIGURE 4-1

"TRENDS" IN THE JOB INFORMATION TEST



Respondent A was in the lower bracket of the trichotomy at T1, respondent B was in the middle bracket at T1, and respondent C was initially in the upper bracket. If there were evidence of a trend-type change across the T1 to T3 interval, we would expect the shift between T2 and T3 to be in the same direction as the observed shift between T1 and T2. The dotted lines in the "Expected T3 Score" column of the Figure indicate that the magnitude of the expected T2 to T3 shift is likely to be less than that observed between T1 and T2. Among the several reasons for this expected decrease in magnitude is the one owing to the crude classification procedure inherent in a trichotomy. That is, rather arbitrary cutting points were used to separate low and middle scores, and middle and high scores. Thus two respondents whose scores differ very slightly may fall on opposite sides of one of these cutting points. Since the observed scores for these two respondents may differ only as a function of measurement error, we may well have misclassified one or both of them in trichotomizing the scores.

The effect of such misclassifications may be seen better if we consider a specific example. Suppose that several respondents whose *true* scores should have placed them initially in the Middle category at T1 had a sufficiently large (and negative) measurement error component in their observed scores that they were classified as Low instead. Now if there really are differences in the way that those correctly classified as Low, Middle and High change across time, the misclassified respondents above ought to behave more like the Middle group than the Low group. Since the Middle group will tend to remain fairly stable across the total interval, these misclassified respondents would not be expected to shift upwards to the same degree as would those respondents correctly classified as Low. Thus, the overall T2 to T3 shift for an A-type respondent (see Figure 4-1) would not be expected to be as large as the shift observed between T1 and T2.

The second column of plots in Figure 4-1 presents the observed T3 scores for respondent Types A, B, and C. As can readily be observed (and as we will shortly see in more statistical terms), no support is found for the notion that trend-like changes are occurring in the Job Information Test. On the contrary, the direction of the T2 to T3 change for respondent Types A and C is exactly opposite to the direction expected if a trend model were to be used to account for the observed shifts.

What we sought at this point was a procedure which would take advantage of all three observations and of whatever evidences of trends could be observed. Our initial efforts were focused on a method which could be used to discover whether the T2 version of the criterion dimension would measurably increase the prediction of the T3 score beyond whatever predictive power the T1 score had. We reasoned that if there were trend-like changes occurring for some respondents, a model based on additive prediction would not fit the data as well as one which permitted prediction from an interactive term. This additional prediction would be due to the fact that a trend-like shift would necessitate a non-additive explanatory term to account for those changers whose scores are deviating from the typically observed overall pattern of stability. This is not to say that the discovery of a non-additive explanatory term would indicate that trends are present; rather, it is to say that no trends are likely to be discovered if an additive model fits the data as well as a model which permits interaction. Thus, the presence of interaction is seen to be a necessary, but not sufficient, condition for the discovery of trend-like changes in situations where the overall stability is high.

The procedure which satisfied the objectives described above involves comparing the additive prediction of a Time 3 criterion dimension from T1 and T2 versions of that dimension with prediction from a specially-constructed combination of the Times 1 and 2 scores. This special T1 and T2 combination score is defined as follows:

1 - Trichotomize the T1 and T2 versions of the criteria via a bracketing procedure which includes in the middle bracket observations falling within 3/4 of a

standard deviation of the overall mean.[2] (If the scores were distributed normally about this mean, then the middle bracket would include about 55% of the scores with the upper and lower categories each containing about 22%.)

2 - Develop a nine-category T1 by T2 composite score for each dimension from the T1 by T2 bivariate table based on the trichotomies from step 1. The nine values were assigned according to the table below.

TABLE 4-1

CONSTRUCTION OF A NINE-CATEGORY COMPOSITE (T1 BY T2) SCORE

Time 1 Trichotomy	Time 2 Trichotomy: 1	2	3
1	1	2	3
2	4	5	6
3	7	8	9

3 - Predict to the T3 criterion from the T1 and T2 trichotomies using Multiple Classification Analysis (MCA).

4 - Predict to the T3 criterion from the T1 by T2 composite score using one way analysis of variance.

---

[2]The "overall mean" is the average of the means of the criteria at all three points in time. For the most part, the standard deviations were equivalent for the three time periods. Where this was not the case, an average of the three standard deviations was used.

5 - Compare the multiple correlation (R) from step 3 with the eta coefficient (η) from step 4.

The comparison in step 5 of this procedure provides the essential data for indicating trends. For if the additive combination of T1 and T2 scores (summarized by the multiple correlation coefficient, R, in the MCA analysis) does as well as the prediction based on the composite score (as reflected by the eta coefficient in the analysis of variance), then there is little support for the notion that there is an overall trend in the T1 to T3 changes. In other words, the direction of the T1 to T2 change should be directly related to the direction of the T2 to T3 change if trends are to be observed.

For example, consider an individual with a composite score of 8 (see Table 4-1). This individual has dropped from the upper to the middle bracket in the T1 to T2 interval. If this drop were indicative of a trend, then those with composite scores of 8 would be expected to have lower T3 scores than would people who have composite scores of 5 (i.e., those whose scores remained stable in the middle category at T1 and T2). As the data below will document, just the opposite effect is observed.

To illustrate the application of this procedure, the Job Information Test will be used as the criterion. This dimension was chosen because it did evidence an overall mean shift upwards across both the Times 1 to 2 and the Times 2 to 3 intervals. (See Table 3-1.) In each of the nine cells of Table 4-2 below, the frequencies (n), means ($\overline{x}$), and standard deviations (sd) of Time 3 scores are presented for the respective combination of Time 1 and Time 2 scores. Additionally, an expected cell mean ($\hat{x}$) based on the adjusted MCA coefficients is given for each cell. Also shown in the table as marginal totals are the frequencies (N) and means ($\overline{X}$) for the T1 and T2 trichotomous variables. Finally, the MCA coefficients for each level of the two trichotomies are displayed in the last row and column, and the multiple correlation coefficient (R) and eta coefficient (η) are given in the lower right corner of the table.

The two coefficients (R and η) are obviously extremely similar, suggesting that the additive prediction (via MCA) of the T3 criterion from the T1 and T2 trichotomies is almost as good as the prediction which also incorporates interaction between the T1 and T2 trichotomies in predicting the T3 criteria. How good is the prediction from the MCA? To answer this question, a cell mean was predicted for each of the 9 cells in the table, using the MCA coefficients. (For example, the prediction for the 1,1 cell was obtained by taking the algebraic sum of the grand mean and the row 1 and column 1 coefficients; $\hat{x}_{1,1} = 19.07 + (-1.04) + (-2.45) = 15.58$.)

TABLE 4-2

JOB INFORMATION: ADDITIVE VS. COMPOSITE PREDICTION

		TIME 2: 1	TIME 2: 2	TIME 2: 3	TOTAL	MCA COEFFI-CIENT
TIME 1	1	n=161 $\bar{x}$=15.4 $\hat{x}$=15.6 sd=3.0	n=215 $\bar{x}$=18.2 $\hat{x}$=18.1 sd=2.7	n=15 $\bar{x}$=19.8 $\hat{x}$=19.7 sd=2.4	N=391 $\bar{X}$=17.1	-1.04
	2	n=81 $\bar{x}$=17.1 $\hat{x}$=16.8 sd=3.2	n=494 $\bar{x}$=19.2 $\hat{x}$=19.3 sd=2.6	n=198 $\bar{x}$=20.9 $\hat{x}$=20.9 sd=2.1	N=773 $\bar{X}$=19.4	.16
	3	n=2 $\bar{x}$=20.0 $\hat{x}$=18.0 sd=1.4	n=84 $\bar{x}$=20.5 $\hat{x}$=20.5 sd=2.5	n=113 $\bar{x}$=22.2 $\hat{x}$=22.2 sd=1.5	N=199 $\bar{X}$=21.5	1.42
TOTAL		N=244 $\bar{X}$=16.0	N=793 $\bar{X}$=19.1	N=326 $\bar{X}$=21.3	N=1363 $\bar{X}$=19.07 SD=3.2	R=.581 η=.584
MCA COEFFICIENT		-2.45	.05	1.71		

Inspection of the table reveals remarkable similarities between the observed and predicted cell means. As a matter of fact, the only cell in which these two means differed by more than .3 is the 3,1 cell in which the observed mean is based on only 2 observations! Also noteworthy in the table is the fact that the respondents who drop in score from T1 to T2 have higher T3 scores (and those who gain in score from T1 to T2 have lower T3 scores) than respondents who had similar T2 scores as did the "movers" but whose scores had remained stable at that level from T1. As mentioned previously, these observations are certainly not consistent with the notion that the observed movement reflects a trend over time.

Another variable which showed a consistent movement in the means across time (see Table 3-1) was Positive School Attitudes. Of special interest was the fact that the attitudes of those staying in the same school were getting consistently less positive as time passed. In Table 4-3 below are displayed data parallel to those given for Job Information in the preceding table. Again, no evidence of trend-like shifts across the Times 1 to 3 interval may be observed.

Nine other criterion dimensions were examined (using the same procedure) for evidences of trends. As for the two dimensions reported above, remarkable similarities between the multiple correlation and eta coefficients were observed. These

TABLE 4-3

POSITIVE SCHOOL ATTITUDES:
ADDITIVE VS. COMPOSITE PREDICTION

		TIME 2: 1	TIME 2: 2	TIME 2: 3	TOTAL	MCA COEFFICIENT
TIME 1	1	n=113 $\bar{x}$=249 $\hat{x}$=248 sd=52	n=90 $\bar{x}$=282 $\hat{x}$=284 sd=45	n=14 $\bar{x}$=321 $\hat{x}$=315 sd=62	N=217 $\bar{X}$=267	-21.20
	2	n=143 $\bar{x}$=268 $\hat{x}$=268 sd=47	n=379 $\bar{x}$=304 $\hat{x}$=304 sd=42	n=135 $\bar{x}$=335 $\hat{x}$=334 sd=44	N=657 $\bar{X}$=302	- 1.40
	3	n=35 $\bar{x}$=282 $\hat{x}$=282 sd=50	n=205 $\bar{x}$=319 $\hat{x}$=317 sd=50	n=213 $\bar{x}$=347 $\hat{x}$=348 sd=44	N=453 $\bar{X}$=329	12.18
TOTAL		N=291 $\bar{X}$=262	N=674 $\bar{X}$=305	N=362 $\bar{X}$=342	N=1327 $\bar{X}$=306 SD=55	R=.541 η=.541
MCA COEFFICIENT		-36.15	-.49	29.97		

relationships are summarized in Table 4-4 below; the more detailed presentation of the data from analyses of these dimensions is given in Appendix C.

TABLE 4-4

OVERALL PREDICTION OF NINE TIME THREE CRITERION SCORES FROM AN ADDITIVE COMBINATION OF TIME ONE AND TIME TWO SCORES (R) VS. PREDICTION FROM A TIME ONE BY TIME TWO COMPOSITE SCORE (η)

Dimension Name	R	η
Negative School Attitudes	.540	.543
Academic Achievement Value	.449	.453
Internal Control	.514	.515
Self-Esteem	.615	.616
Negative Affective States	.646	.647
Social Values	.558	.558
Ambitious Job Attitudes	.502	.504
Aspired Occupational Status	.645	.652
Delinquent Behaviors	.640	.641

In short, the procedure outlined earlier in the chapter uncovered no evidence of trend-like changes in the eleven dimensions examined. Thus, the procedure adds little if anything to change the overall picture of stability which has previously emerged. However, the detailed tables (4-2, 4-3, and Appendix C) document the nature of the shifts which are taking place from beginning sophomore to end junior to end senior years for those boys who did stay in the same school during this entire period. For this descriptive purpose, these tables may have some utility. Let us now turn our attention to the matter of subgroup analyses, the second question to be addressed in this chapter.

Analysis of Subgroups

The fact that few overall trends have been observed thus far in no way precludes the possibility that subgroups within the total sample may be changing in identifiable and interesting ways. For example, a monograph has recently been written by other members of the Youth in Transition staff which focuses attention on comparisons between and among three major subgroups: those who drop out of high school, those who graduate from high school but do not continue their education further, and those who pursue their education beyond high school (Bachman, et al., 1971). Subsequent analysis efforts to be reported in later publications will be aimed at other subgroups of interest; examples of such subgroups include respondents with military experience, graduates of work-study programs, and those who continue their formal education beyond high school.

One general method of determining whether or not subgroups differ along the criterion dimension is to develop a set of subgroup categories which are both totally inclusive and mutually exclusive. Having thus classified everyone in the sample using this "subgroup variable," two MCA runs are made to predict a dependent variable of interest; one run simply predicts from the set of selected independent variables to the dependent variable. The other run is similar except that the "subgroup variable" is added to the predictor list. Should the multiple correlation coefficient from the second run be significantly higher than that from the first run, one can conclude that there are differences among the subgroups which are worth further exploration. However, should the two multiple correlation coefficients be essentially the same, then one can conclude that the "subgroup variable" does not contribute much to the predictive ability of the set of independent variables.

In other instances, one's interest in subgroups extends to more complex analytic areas, however. Within the observed general setting of overall stability, it would be necessary for some subgroups to be moving in one direction while other subgroups are changing in the opposite direction on the same dimension, and detection of such "counterbalancing" would require a different technique from that described above. If we should observe instances where definite but opposing shifts could be identified with various subgroups which produced no observable effect in the aggregate, then we would have identified an area where more sophisticated analyses efforts would be required. In such areas, for example, we might want to do regression analyses within each subgroup using the same set of predictor variables. Of particular interest in such analyses would be the identification of any variable whose effect is facilitative for one or more subgroup(s) and debilitative for other(s).

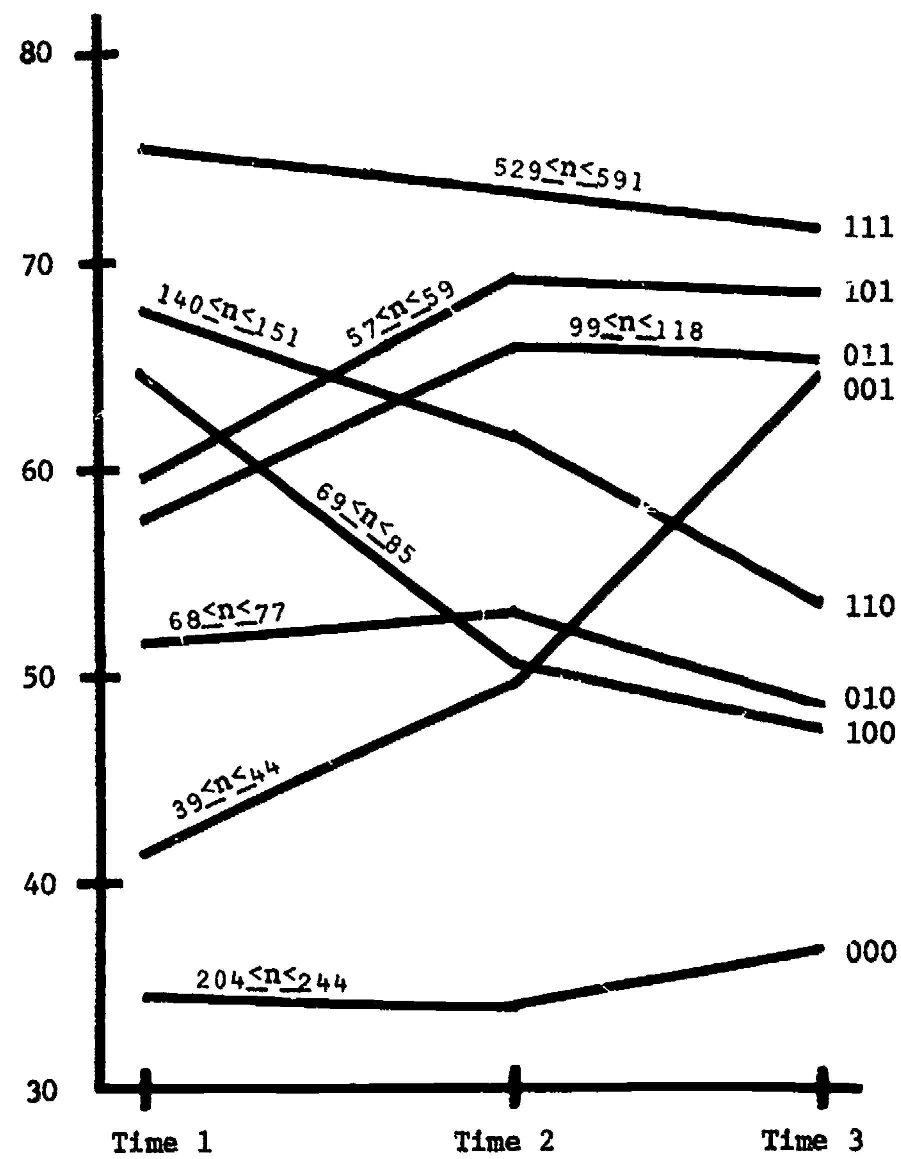
At the present time, analysis efforts of this type have not been undertaken for several reasons:

(1) subgroups have only recently been identified, and in many cases even initial analyses are not yet underway,
(2) considerable effort to date has been focused on the investigation of family background predictors whose effects seem to be rather unidirectional,
(3) school effects analyses (perhaps the most interesting area for investigations of this type) are still in the planning stages due to the very complex and time-consuming nature of developing measures of characterizing the schools, and
(4) as mentioned earlier, we need to spend most of our time developing and utilizing procedures which have general utility across most or all areas of our analytic framework; thus, concentrating efforts on one or more subgroups has been a tempting activity which we have had to avoid.

In order to anticipate what such analyses might tell us, the author has conducted a brief investigation of the relationship between College Plans and Status of Aspired Occupation. Of special interest here is the question of what effects, if any, on a boy's Aspired Occupation result from a change in his College Plans. The data displayed in Figure 4-2 are of interest for several reasons. First of all, we note that those who consistently plan to go to college (the 111 group) have the highest aspired occupational status at all three points in time, whereas those who never plan to go to college (the 000 group) have the lowest occupational status across time. Secondly, both at Time 1 and Time 3 (and also at Time 2 except for the 101 group) those planning to attend college have higher aspired occupational status than do those not planning to attend college. Thirdly, a shift in college plans is accompanied (in all cases except for the 101 group) by a similar shift in aspired occupational status; that is, the two groups who originally planned college but subsequently dropped those plans (110 and 100) are observed to have the largest drop in aspired occupational status at the time they dropped their college plans, whereas those groups not originally planning college who later changed their plans (011 and 001) may be seen to have the largest gain in aspired occupational status at the time their plans changed to include college.

FIGURE 4-2

STATUS OF ASPIRED OCCUPATION VS. COLLEGE/NON-COLLEGE PLANS



The ranges represent the minimum and maximum frequencies underlying the points along each line. The triads in the right margin identify whether respondents represented by each line had college plans (1) or did not have college plans (0) at Times 1, 2, and 3 respectively.

The data in Figure 4-2 provide an illustration of one type of analyses which can be conducted for exclusive subgroups (in this case the college-bound vs. the noncollege-bound). More sophisticated analyses of such subgroups will undoubtedly be undertaken as a part of forthcoming publications.

Identification of Subgroups of "Changers"

Subgroups described thus far have been identified because of conceptual or substantive interest. It is also possible to use empirical procedures to identify subgroups of "changers" for additional analyses.

For example, Trent identified three groups by defining an "exceptional" change group as those whose (raw) change scores were three-quarters of a standard deviation or more above the average change and a "negative" change group whose change scores were at least three-fourths of a standard deviation below the average, with the remaining group members falling in the "average" change group (Trent and Medsker, 1968, pp. 178-218). Trent and Medsker spend considerable time analyzing differences among these three change groups. Of particular interest to our present discussion are the data presented in Table 4-5 below.

Now, completely apart from whatever "real" changes are taking place, we would expect those with the lowest initial scores to have higher final

TABLE 4-5[3]

INITIAL AND FINAL SOCIAL MATURITY
MEAN SCORES FOR THREE CHANGE GROUPS

	Exceptional Changers	Average Changers	Negative Changers
Initial Score	47.81	50.46	55.33
Final Score	66.82	56.21	48.67

scores (relative to those in the group), and we would also expect those with the highest initial scores to have relatively lower final scores. Again, this may reflect nothing other than the fact that measurement error "artificially" depressed the observed initial low scores and inflated the observed initial high scores. Since this measurement error is assumed to be uncorrelated across time, roughly half of those in the extreme groups will have a final score measurement error component which is opposite to the initial score measurement error component, and we will thus observe that those with extreme initial scores will have less extreme final scores (i.e., their scores will regress toward the mean). From the data presented by Trent, it is not possible to know how much of the observed change might be due merely to regression; however, it is at least safe to conclude that whatever regression effect

[3]This table is based on Table 56, p. 189, in Trent and Medsker, 1968.

is present has been confounded with the changes observed. For this reason, the present author does not find such empirical identification of subgroups of "changers" to be particularly helpful.

It is conceivable that a residualized gain score might be useful in empirically identifying subgroups for subsequent analyses. The author is quite doubtful about the probability that the considerable effort necessary to produce such residualized gain scores would be warranted in the present data for this purpose alone. Had earlier analyses suggested that for more basic purposes such a gain score was useful, we would have doubtless investigated its utility in these analyses as well.

Summary

The first part of this chapter was devoted to the description of a procedure designed to investigate trend-like changes in the criterion dimensions. Applying this procedure to eleven dimensions resulted in no evidence of trends in the Times 1 to 2 vs. the Times 2 to 3 changes.

The second part of the chapter contained a brief description of some analyses performed on conceptually-defined subgroups. An illustration of one type of data display was given which showed how shifts into and out of the college-bound subgroup are accompanied by corresponding shifts in the boys' aspired occupational status. Finally, a limitation (due largely to regression effects) in the empirical identification of subgroups of "changers" was described and illustrated, using data from a previous study of post-high-school youth.

Chapter 5

SUMMARY OF THE PROPOSED STRATEGY

The previous four chapters have described the evolution of a strategy for longitudinal analyses of survey panel data. In this final chapter, the proposed analytic model will be reviewed briefly, with special attention given to some critical questions around which application of the model is based.

For What Kinds of Studies Is the Model Intended?

The proposed analytic strategy is focused on longitudinal analysis of panel data. Specifically, the dependent or criterion variables of interest are assumed to be measured on the same set of respondents at two or more points in time. Since many of the statistical techniques employed in the application of the model require relatively large sample sizes,[1] it is assumed that the major use of the strategy will be found in survey panel studies. However, except for considerations such as those just described, the model should find applicability in any panel study.

---

[1]This requirement is necessary in order to get relatively small sampling errors for the statistical estimates employed in techniques such as MCA.

What Kind of Change Score(s) Should be Used?

An investigator must first decide what are his intended uses for change scores. (A partial list may be found on pages 17-18.) If one of the purposes of a change score is to identify for further study those individuals who have gained or lost an exceptional amount during the interval between observations, a form of residualized gain score (see Cronbach and Furby, 1970, pp. 77-80) might be helpful. Except for this rather unique purpose, however, the calculation of any type of gain score appears to be unnecessary and, as we saw in Chapter 3, sometimes misleading.

An additional limitation in the use of change scores is that they utilize data from only two points in time. In pre-post designs, this is not a serious problem; but in panel designs which employ more than two observations, a host of problems arises. As Figure 1-2 indicated, a "four-wave" panel study provides that change could be studied across six different intervals. How one chooses only some intervals for examining change (thereby eliminating others) is a difficult problem. This is especially true if the rates of change differ from one interval to another. In short the author finds himself in agreement with Cronbach and Furby (1970, pp. 77-80) who state that

> ...gain scores are rarely useful, no matter how they may be adjusted or refined (Cronbach and Furby, 1970, p. 68).

Description of a "Parallel Prediction" Approach

The proposed analytic model utilizes the repeatedly measured dependent variables (criteria) as static scores. Specifically, it proposes to make separate predictions from a specified set of independent variables to the First, Second, Third,..., and Nth criterion scores.[2] Of special interest in these analyses is the identification of criterion variables whose overall predictability is changing meaningfully across time. Also of interest at this stage is the relative importance of the independent (predictor) variables in the multiple prediction equations. Specifically, predictors which systematically increase (or decrease) in explanatory power across time are deserving of further attention.

In cases where neither the overall predictability nor the relative explanatory power of the predictor variables change over time, the multiple prediction equations may be of considerable interest in their own right. In such cases, the regression equations in the prediction of separate static criteria will resemble one another quite closely, and any one of them could be used to describe the relationships that exist between the criterion and the set of predictors.

In instances where the relative power of the set of predictors does change across time, one may

---

[2]Multiple regression models (both linear and MCA) will be useful at this stage in the analysis.

be interested in knowing whether there is consistency in the kinds of changes which are taking place. A procedure is described in Chapter 4 which may be used for identifying such "trends."

Summary

A "parallel prediction" model for longitudinal analysis has been described. The model makes separate use of each repetition of the criterion dimension. The proposed strategy seems to be widely applicable in studies employing panel designs; it avoids the messy philosophical and analytical problems inherent in the use of any and all kinds of change scores, and it provides descriptive data which are interesting in their own right.

Chapter 6
EPILOGUE

Since completing the first five chapters (which were submitted in partial fulfillment of the degree of Doctor of Philosophy in the University of Michigan)[1], the author has applied the proposed parallel prediction strategy in a limited set of analyses of Youth in Transition data. This epilogue will be devoted to a brief report of the results of these analytic efforts.

The Use of Raw Change Scores to Identify Differential Shifts in Subgroups Across Time

In Chapter 4 (see esp. pp. 63-68), a procedure was proposed for investigating the issue of change via examining subgroup shifts across time. Reported below are results from two sets of such analyses, each set predicting to six important criterion dimensions. The first set is based on subgroups identified by the Socio-Economic Level(SEL) of the respondent's family, and the second set on the respondent's own level of intelligence, as measured by the Quick Test (QT). These two variables have been

[1]Jerald G. Bachman and William M. Cave served as Co-Chairmen. Other committee members were Frank M. Andrews and LaVerne S. Collet.

chosen both for their theoretical interest as well as for their predictive utility with Youth in Transition criteria.[2]

First of all, let us look at the cross-time relationships between SEL and Self Esteem. In Chapter Three we saw that (1) Self Esteem scores were quite stable in their means and standard deviations across time (Table 3-1), (2) the autocorrelations were inversely related to the length of the interval between observations (Table 3-2), (3) the internal consistency of the scale remained stable at a respectable level from Time One to Time Three (Table 3-3), and (4) the reliabilities of the raw and regressed gain scores were not too impressive (Table 3-4). Therefore, it is not too surprising to note in Table 6-1 below that the relationship between SEL and Self Esteem does not deteriorate to any degree across time and that the overall raw change (Time 3-Time 1 static scores) in Self Esteem is rather unrelated to SEL. Similar observations are obtained for Negative Affective States and Occupational Aspirations. However, two of the other three criterion dimensions (Academic Achievement Value and Social Values) demonstrate relationships between overall change and SEL which are larger than the static relationships. Can we observe interesting subgroup shifts in these latter two cases?

[2]See Bachman, 1970, for operational definitions of these variables and for empirical evidence of their predictive power.

TABLE 6-1

RELATIONSHIPS BETWEEN SEL AND
SIX SELECTED CRITERIA ACROSS TIME[1]

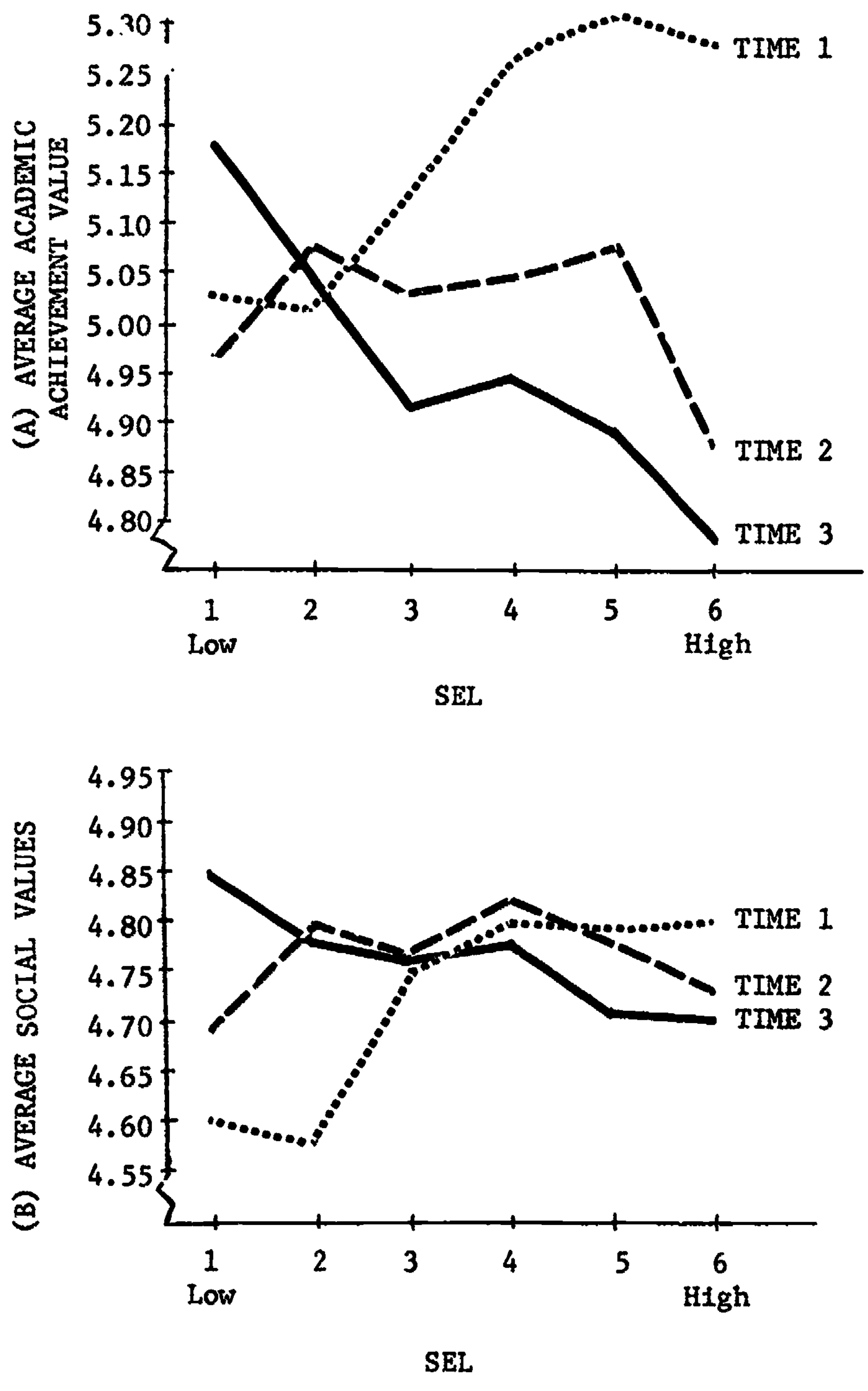
Criterion Dimension	Time 1 (Static) Score	Time 2 (Static) Score	Time 3 (Static) Score	Time 1 to 3 (Raw Change) Score
Self Esteem	.13	.10	.11	.04
Negative Affective States	.08	.04	.08	.06
Occupational Aspirations	.35	.29	.32	.05
Ambitious Job Attitudes	.19	.11	.12	.10
Academic Achievement Value	.15	.08	.12	.19
Social Values	.15	.07	.08	.17

[1]Entries in this table are eta coefficients.

In order to answer this question, let us first plot the results of our "parallel predictions" from SEL to the static scores. (See Figure 6-1 below.) As can be seen in Part A of this figure, the lowest SEL group tended to value Academic Achievement more in their senior year than in their sophomore year, whereas the groups who are average or above average in SEL decreased in their value of academic achieve-

FIGURE 6-1

SEL VS. AVERAGE STATIC (A) ACADEMIC ACHIEVEMENT VALUE AND (B) SOCIAL VALUES SCORES AT TIMES 1, 2, AND 3



ment. There is a rather strong monotonic relationship between SEL and average difference in Academic Achievement in these data. Thus, the relationship between SEL and raw change on this dimension observed in Table 6-1 was indicative of a very interesting subgroup shift. It should be pointed out that this relationship was observed in spite of the fact that the magnitude of the static relationships (as reflected by eta coefficients) did not change dramatically across time.

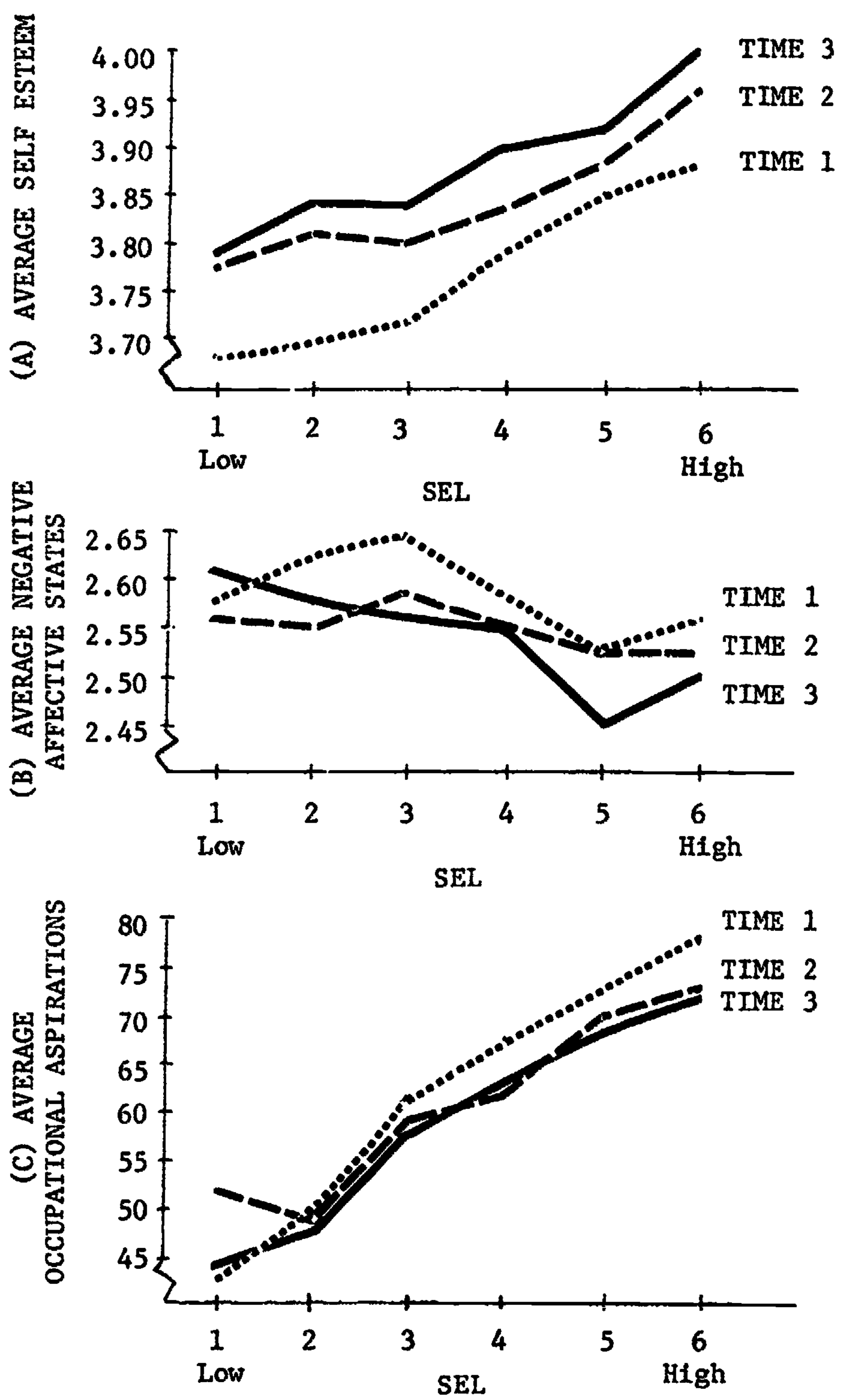
In Part B of Figure 6-1, we may examine the data relating SEL to Social Values scores across time. Here we see that the relationship between SEL and Social Values has dropped markedly from Time 1 to Time 3. However, we again observe a rather strong monotonic relationship between SEL and the average group differences between Times 1 and 3. As in the previous case, interesting subgroup shifts were evidenced in the figure in a situation where the raw change score indicated such a relationship.

In contrast to these situations, let us examine similar plots for the three criteria for which SEL and overall change were unrelated. (See Figure 6-2 below.) In each of these figures, the three lines are observed to be relatively parallel to one another; thus the six SEL subgroups do not appear to be changing differentially in these situations where the raw change score indicated little, if any, relationship.

The criterion dimension of Ambitious Job Attitudes is of interest because it provides a situation where the static relationship with SEL decreases

FIGURE 6-2

SEL VS. AVERAGE STATIC SCORES AT TIMES 1, 2, AND 3 FOR (A) SELF ESTEEM, (B) NEGATIVE AFFECTIVE STATES, AND (C) OCCUPATIONAL ASPIRATIONS

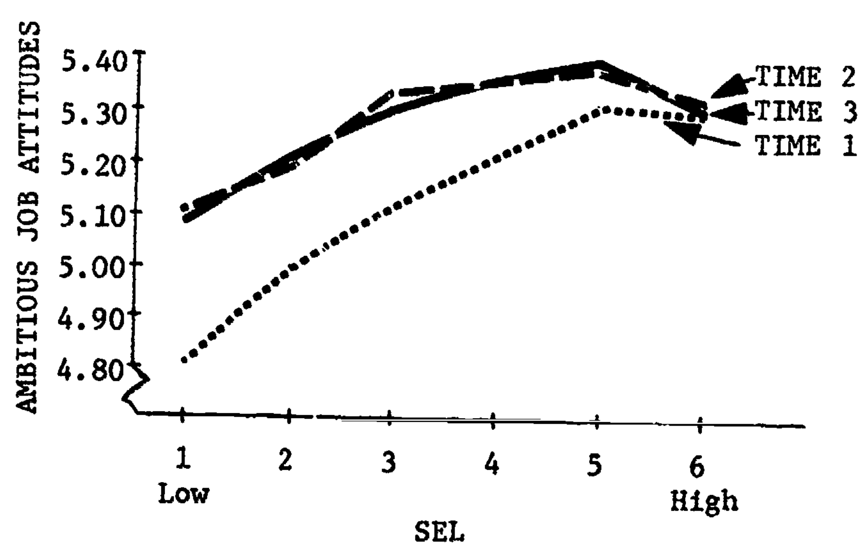


across time (as we saw with Social Values earlier), but where the relationship between SEL and overall change is considerably smaller than was the case with Social Values. Does this suggest that only some of the subgroups are shifting across time?

To answer this question, examine Figure 6-3. It may be noted here that there is a largely monotonic relationship between SEL and the average subgroup differences between Time 1 and Time 3 on Ambitious Job Attitudes, but that this relationship is a good deal stronger for above average SEL groups than for below average. Thus, it does appear to be the case that about half of the six SEL subgroups are accounting for most of the SEL vs. raw change relationship.

FIGURE 6-3

SEL VS. AVERAGE STATIC AMBITIOUS JOB ATTITUDES SCORES AT TIMES 1, 2, AND 3



Previous analyses have documented the fact that SEL and various measures of intelligence are strongly related in our sample of boys.[3] Therefore, it should come as no surprise to learn that the relationships between one such intelligence measure, the Quick Test (QT), and the static and raw change scores on the six selected criterion dimensions are very similar to

TABLE 6-2

RELATIONSHIPS BETWEEN QT AND SIX SELECTED CRITERIA ACROSS TIME[1]

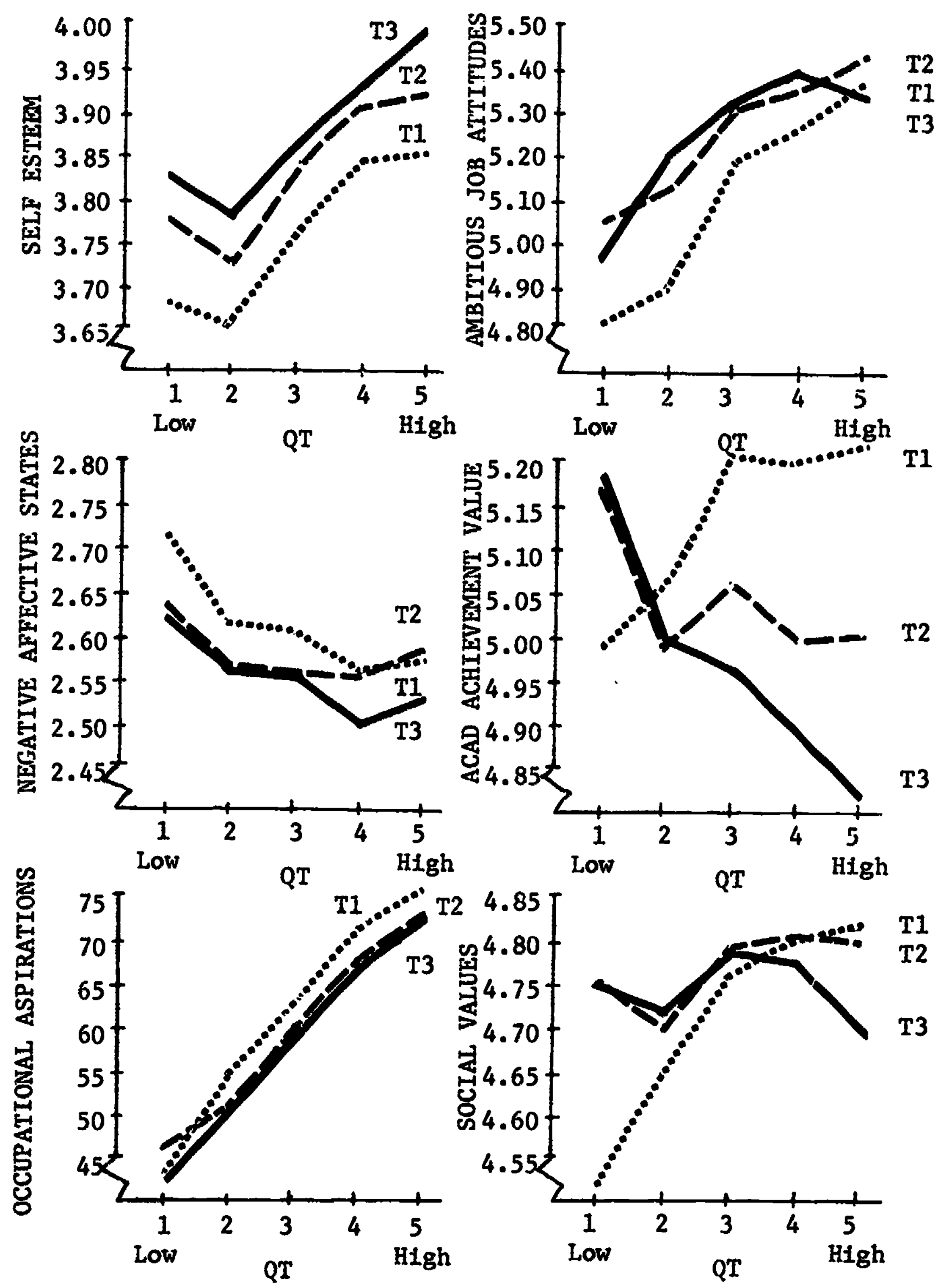
Criterion Dimension	Time 1 (Static) Score	Time 2 (Static) Score	Time 3 (Static) Score	Time 1 to 3 (Raw Change) Score
Self Esteem	.14	.13	.12	.05
Negative Affective States	.06	.04	.06	.02
Occupational Aspirations	.33	.31	.33	.02
Ambitious Job Attitudes	.23	.16	.16	.11
Academic Achievement Value	.11	.07	.11	.18
Social Values	.14	.08	.06	.13

[1]Entries in this table are eta coefficients.

[3]See Bachman, 1970, for several indications and discussions of these relationships.

FIGURE 6-4

QT VS. AVERAGE STATIC SCORES FOR SIX CRITERIA
AT TIMES 1 (•••••), 2 (—), and 3 (━━)



those already reported for SEL. Data summarizing these relationships (in a fashion similar to the treatment afforded SEL) may be found in Table 6-2 and Figure 6-4. Because of the great similarity to the SEL displays, these data will not be discussed here. The interested reader is invited to explore them as he wishes.

From these few analyses, it would appear as if the proposed parallel prediction model may be a good one for examining differential shifts among subgroups. We have seen evidence that even when the criterion dimension possesses considerable stability across time, the procedure may provide data of considerable interest. Furthermore, the relationships with the overall raw change scores appear to be useful indicators of situations in which subgroups are shifting differentially across time.

Application of the Proposed Procedure to Groups of Empirically-Defined "Changers"

In Chapter 4, we examined a technique used by two previous researchers (Trent and Medsker, op. cit.) to identify subgroups of interest by using the raw change scores themselves. The procedure they suggested for this purpose is based on raw change scores, and it results in three subgroups of potential interest: Exceptional Changers (EC) are those who change most positively, Negative Changers (NC) are those who change most negatively, and Average Changers (AC) are those who change the least in either direction.[4] As we noted in Chapter 4, this kind of definition of change does not consider whatever regression effect may be observed in the data, and because of this fact, the utility of the procedure seemed to be questionable, at best.

In order to consider further the utility of this type of analysis, the author has defined the three types of change groups just described for each of the six selected criterion dimensions reported in the preceeding section. For each criterion separately, average Time 1 and Time 3 scores were then calculated for each type of change group. Results of these analyses are reported in Table 6-3. By comparing the Time 1 and Time 3 subgroup means to the Time 1 and Time 3 grand means, we may observe the following:

(1) EC groups always have average scores which, at Time 1, are below the (Time 1) grand mean and, at Time 3, are above the (Time 3) grand mean.

(2) NC groups always have average scores which, at Time 1, are above the (Time 1) grand mean and, at Time 3, are below the (Time 3) grand mean.

(3) AC groups always have average scores which are very near the grand mean, both for Times 1 and 3.

In short, the lower the initial score (or the higher the final score), the more likely one is to be an Exceptional Changer, and the higher the initial score

---

[4] See p. 68 for a review of the specific procedure, if desired.

TABLE 6-3

"CHANGE GROUPS" VS. AVERAGE STATIC SCORES AT TIMES 1, 2, AND 3

Criterion Dimension		Exceptional Changers	Average Changers	Negative Changers	Grand Mean
Self Esteem	T1	3.41	3.80	4.09	3.77
	T2	3.88	3.85	3.78	3.84
	T3	4.17	3.90	3.55	3.88
no. of cases		335	708	319	1362
Negative Affective States	T1	2.30	2.57	2.98	2.59
	T2	2.69	2.53	2.48	2.56
	T3	2.94	2.49	2.22	2.53
no. of cases		282	799	264	1345
Occupational Aspirations	T1	41.7	67.7	78.9	65.4
	T2	60.8	65.8	55.5	63.0
	T3	72.6	64.5	39.4	61.2
no. of cases		157	601	170	928
Ambitious Job Attitudes	T1	4.56	5.19	5.65	5.15
	T2	5.29	5.32	5.26	5.30
	T3	5.73	5.34	4.81	5.31
no. of cases		272	794	269	1335
Academic Achievement Value	T1	4.49	5.28	5.62	5.19
	T2	5.03	5.08	4.94	5.04
	T3	5.42	5.04	4.29	4.94
no. of cases		265	740	303	1308
Social Values	T1	4.18	4.77	5.20	4.75
	T2	4.74	4.78	4.81	4.78
	T3	4.98	4.78	4.50	4.76
no. of cases		248	808	284	1340

(or the lower the final score), the more likely one is to be a Negative Changer. Thus, the disturbing relationships (between type of changer and initial and final score) reported earlier in Table 4-5 are most assuredly a result of defining change groups in this fashion.

Summary

The proposed parallel prediction model has been applied to subgroup analyses. Plotting the average criterion value separately at each point in time for each subgroup provides a concise picture of subgroup shifts across time. Even when means and standard deviations of the criterion were observed to be quite stable from sophomore through senior years, interesting subgroup shifts have been observed. Of particular note is the fact that relationships between raw change and subgroup level provided consistent indicators of differential subgroup shifts.

Additional analyses of the type described earlier aimed at identifying three subgroups using raw change scores further documented the confounding of change with regression inherent in this definition. Those who were defined as having the largest positive change were consistently observed to have the lowest average initial score and the highest average final score. Similarly, those who were defined as having the largest negative change were consistently observed to have the highest average initial score and the lowest average final score. The value of analyses based on groups defined in this fashion seems doubtful, at best.

The early identification of subgroups is thus seen to have a facilitating effect in longitudinal analyses. Examining the pattern of subgroup shifts across time may provide interesting and insightful looks at the data, even if no aggregate shifts are observed. In spite of their well-documented weaknesses, raw change scores may greatly facilitate such analytic efforts, provided they are interpreted with caution based on a clear realization of their potential for bias.

APPENDICES

Appendix A

THE APPLICABILITY OF THE COLEMAN MODEL FOR ESTIMATING TRUE GAIN FROM THREE WAVES OF DATA

The general model given by Coleman is an expression for the rate of change $(\frac{dx_1}{dt})$ in a variable as a linear additive function of the variable itself (X_1) and other independent variables $(X_2, X_3, \ldots, X_N)$. For the sake of simplicity, we will examine the situation where there is only one independent variable (X_2) but the discussion may be readily extended to the general case of N independent variables.

This leads to a differential equation as follows:

$$\frac{dx_1}{dt} = a + b_1X_1 + b_2X_2.$$

Integration of this expression yields an equation of the following form:

$$X_{1t} = \frac{a}{b_1}(e^{b_1\Delta t}-1) + e^{b_1\Delta t}X_{10} + \frac{b_2}{b_1}(e^{b_1\Delta t}-1)X_2.$$

Here, the second subscript on the X_1 terms indicate the source of the measure; the initial time is represented by 0 and some later time is represented

[1]Coleman, 1968, pp. 28-478. Formula 11.66 (p. 456) is incorrect (Coleman, personal communications, 1969). It should be:

$$b_{1e} = \ln\frac{(b_1{}^*)^2}{b_1{}^\circ}.$$

by t. Note that this is a linear equation of the form: $X_{1t} = A^* + b_1^* X_{10} + b_2^* X_2$. Linear regression may be used to estimate the values of the coefficients in this equation. If we set $C^* = \frac{b_1 \Delta t}{e^{b_1 \Delta t} - 1} = \ln \frac{b_1^*}{b_1^* - 1}$, then the coefficients in the original differential equation are:

$$a = \frac{a^* c^*}{\Delta t},$$

$$b_1 = \frac{\ln b_1^*}{\Delta t}, \text{ and}$$

$$b_2 = \frac{b_2^* C^*}{\Delta t}.$$

Coleman next discussed the effect of measurement error in X_1 on the estimates. Rather than introducing the mathematics here, the more important thing is to understand that we are trying to partition observed or raw change into a component reflecting true change and a second component due to measurement error. This partitioning is extremely important to avoid the situation in which "...measurement error is masquerading as change" (Coleman, 1968, p. 453) by causing changes in the observed scores over time when there is no true change taking place. To this end data from the third observation are brought into the picture. Now, if only measurement error is causing the observed change in the dependent variable then the relation between the Time 1 and Time 2 scores ought to be the same as the relation between the Time 1 and Time 3 scores (or, for that matter, between the Time 2 and Time 3 scores). But if there is no measurement error, the relation between the Time 1 and Time 3 scores should be less than between the Time 1 and Time 2 (or Time 2 and Time 3) scores, because the greater length of time has allowed more change to occur. Coleman proceeds to extend the model in such a way as to permit estimating the relative importance of these two components (Coleman, 1968, pp. 453-456).

The approach appears well suited to our purposes; it encompasses both the issues of unreliability and the simultaneous use of the data from all three observation periods. However, perhaps the most critical assumption underlying the model is subject to question in our study; that is, it is difficult to imagine that the rate of change through the interval spanned by the study is a constant for many of our dependent variables. Because we feel very uncomfortable about making this assumption for many of our variables, the utility of the Coleman model is at best limited.

Appendix B

RELATIONSHIP BETWEEN INITIAL SCORE AND RAW GAIN SCORE

Following are calculations showing that, in the typical case where initial and final scores have approximately equal variances, raw gain scores will show negative correlations with initial scores.

Let A= the initial raw scores, and
let B= the final raw score, and
let G= the raw gain score = B-A.

Now the correlation between the initial and gain scores may be represented by the formula:[1]

$$r_{AG} = \frac{r_{AB}\sigma_B - \sigma_A}{\sigma_G}. \qquad (1)$$

We can express σ_G in terms of A and B as follows:[2]

$$\sigma_G{}^2 = \sigma_{B-A}{}^2 = (1)^2\sigma_B{}^2 + (-1)^2\sigma_A{}^2 = \sigma_B{}^2 + \sigma_A{}^2.$$

Hence $\sigma_G = \sqrt{\sigma_B{}^2 + \sigma_A{}^2}$. (2)

(Note: $\sigma_G = \sigma_A\sqrt{2}$ when $\sigma_B = \sigma_A$.) Now we want to examine the value of r_{AG} in the situation where $\sigma_A \approx \sigma_B$, the usual case. Substituting σ_A for σ_B in equations (1) and (2)

---

[1] Shaycoft, 1967, p. 3-12.

[2] Hays, 1963, p. 236. This formula assumes that A and B are independent.

we get:[3] $r_{AG} = \frac{r_{AB}\sigma_A - \sigma_A}{\sigma_A\sqrt{2}} = \frac{(r_{AB}-1)\sigma_A}{\sigma_A\sqrt{2}} = \frac{r_{AB}-1}{\sqrt{2}}.$

Note that the numerator of this coefficient will always be negative, except when $r_{AB} = 1$ in which case the ratio assumes the value zero. Thus, in the usual case where the variances of the initial and final scores are approximately equal, there will be a negative correlation between the initial score and raw gain score.

[3]Shaycoft's formula (2) on p. 3-12 appears to be in error in this regard.

Appendix C

COMPARISONS OF ADDITIVE AND COMPOSITE PREDICTIONS OF NINE TIME 3 CRITERION SCORES FROM TIME 1 AND TIME 2 SCORES ON THE SAME DIMENSION

Following are tables summarizing the additive and composite predictions of the Time 3 measure from the Time 1 and Time 2 measures for nine criterion dimensions. The entries in these tables follow the format used in Tables 4-2 and 4-3 in the text. Namely, the frequency, mean, and standard deviation of the Time 3 score are given for each of the nine cells. The marginal means and frequencies are given for each value of the Time 1 and Time 2 trichotomies. For the table as a whole, the overall Time 3 mean, standard deviation, and total frequency is given in the cell in the lower right corner. Finally, the adjusted multiple correlation coefficient (from the additive MCA prediction of the Time 3 from the Time 1 and Time 2 scores) and the eta coefficient (from the analysis of variance prediction of the Time 3 from the Time 1 by Time 2 composite variable) are given under the standard deviation in the lower right corner cell.

SELF-ESTEEM:
ADDITIVE VS. COMPOSITE PREDICTION

		TIME 2			TOTAL
		1	2	3	
TIME 1	1	n=163 $\bar{x}$=335 sd=45	n=179 $\bar{x}$=373 sd=41	n=20 $\bar{x}$=419 sd=45	N=362 $\bar{X}$=358
	2	n=119 $\bar{x}$=352 sd=40	n=485 $\bar{x}$=392 sd=39	n=133 $\bar{x}$=426 sd=37	N=737 $\bar{X}$=391
	3	n=8 $\bar{x}$=376 sd=37	n=120 $\bar{x}$=409 sd=37	n=122 $\bar{x}$=440 sd=33	N=250 $\bar{X}$=423
TOTAL		N=290 $\bar{X}$=343	N=784 $\bar{X}$=390	N=275 $\bar{X}$=432	N=1349 $\bar{X}$=388 SD=49 R=.613 η=.616

NEGATIVE AFFECTIVE STATES:
ADDITIVE VS. COMPOSITE PREDICTION

		TIME 2			
		1	2	3	TOTAL
TIME 1	1	n=141 $\bar{x}$=201 sd=46	n=100 $\bar{x}$=239 sd=40	n=9 $\bar{x}$=281 sd=48	N=250 $\bar{X}$=219
	2	n=144 $\bar{x}$=211 sd=36	n=533 $\bar{x}$=251 sd=39	n=110 $\bar{x}$=290 sd=44	N=787 $\bar{X}$=249
	3	n=15 $\bar{x}$=229 sd=55	n=127 $\bar{x}$=270 sd=39	n=150 $\bar{x}$=322 sd=44	N=292 $\bar{X}$=295
TOTAL		N=300 $\bar{X}$=207	N=760 $\bar{X}$=253	N=269 $\bar{X}$=308	N=1329 $\bar{X}$=254 SD=53 R=.644 η=.647

SOCIAL VALUES:
ADDITIVE VS. COMPOSITE PREDICTION

		TIME 2			TOTAL
		1	2	3	
TIME 1	1	n=150 $\bar{x}$=436 sd=41	n=143 $\bar{x}$=463 sd=40	n=26 $\bar{x}$=494 sd=50	N=319 $\bar{X}$=453
	2	n=97 $\bar{x}$=446 sd=38	n=440 $\bar{x}$=471 sd=33	n=130 $\bar{x}$=506 sd=37	N=667 $\bar{X}$=474
	3	n=30 $\bar{x}$=468 sd=68	n=135 $\bar{x}$=486 sd=32	n=168 $\bar{x}$=520 sd=36	N=333 $\bar{X}$=502
TOTAL		N=277 $\bar{X}$=443	N=718 $\bar{X}$=472	N=324 $\bar{X}$=512	N=1319 $\bar{X}$=476 SD=63 R=.557 η=.558

AMBITIOUS JOB ATTITUDES:
ADDITIVE VS. COMPOSITE PREDICTION

		TIME 2			
		1	2	3	TOTAL
TIME 1	1	n=138 $\bar{x}$=478 sd=55	n=202 $\bar{x}$=509 sd=59	n=30 $\bar{x}$=571 sd=58	N=370 $\bar{X}$=503
	2	n=113 $\bar{x}$=490 sd=57	n=413 $\bar{x}$=534 sd=51	n=173 $\bar{x}$=567 sd=49	N=699 $\bar{X}$=535
	3	n=10 $\bar{x}$=510 sd=69	n=114 $\bar{x}$=543 sd=58	n=132 $\bar{x}$=581 sd=53	N=256 $\bar{X}$=562
TOTAL		N=261 $\bar{X}$=484	N=729 $\bar{X}$=529	N=335 $\bar{X}$=573	N=1325 $\bar{X}$=531 SD=62 R=.499 η=.504

ASPIRED OCCUPATION:
ADDITIVE VS. COMPOSITE PREDICTION

		TIME 2			
		1	2	3	TOTAL
TIME 1	1	n=86 $\bar{x}$=29 sd=19	n=34 $\bar{x}$=49 sd=27	n=7 $\bar{x}$=47 sd=21	N=127 $\bar{X}$=35
	2	n=33 $\bar{x}$=38 sd=22	n=304 $\bar{x}$=62 sd=18	n=65 $\bar{x}$=76 sd=16	N=402 $\bar{X}$=63
	3	n=12 $\bar{x}$=52 sd=28	n=105 $\bar{x}$=62 sd=20	n=158 $\bar{x}$=81 sd=14	N=275 $\bar{X}$=73
TOTAL		N=131 $\bar{X}$=33	N=443 $\bar{X}$=61	N=230 $\bar{X}$=79	N=804 $\bar{X}$=62 SD=24 R=.642 η=.652

TOTAL DELINQUENCY:
ADDITIVE VS. COMPOSITE PREDICTION

		TIME 2: 1	TIME 2: 2	TIME 2: 3	TOTAL
TIME 1	1	n=186 $\bar{x}$=119 sd=15	n=113 $\bar{x}$=142 sd=29	n=8 $\bar{x}$=189 sd=61	N=307 $\bar{X}$=129
TIME 1	2	n=141 $\bar{x}$=129 sd=20	n=536 $\bar{x}$=157 sd=32	n=91 $\bar{x}$=196 sd=46	N=768 $\bar{X}$=156
TIME 1	3	n=11 $\bar{x}$=143 sd=23	n=121 $\bar{x}$=175 sd=37	n=120 $\bar{x}$=207 sd=38	N=252 $\bar{X}$=189
TOTAL		N=338 $\bar{X}$=124	N=770 $\bar{X}$=158	N=219 $\bar{X}$=202	N=1327 $\bar{X}$=156 SD=41 R=.638 η=.641

ACADEMIC ACHIEVEMENT VALUE:
ADDITIVE VS. COMPOSITE PREDICTION

		TIME 2			
		1	2	3	TOTAL
TIME 1	1	n=121 $\bar{x}$=434 sd=68	n=91 $\bar{x}$=488 sd=55	n=22 $\bar{x}$=525 sd=79	N=234 $\bar{X}$=464
	2	n=162 $\bar{x}$=461 sd=67	n=340 $\bar{x}$=494 sd=59	n=114 $\bar{x}$=530 sd=60	N=616 $\bar{X}$=492
	3	n=56 $\bar{x}$=473 sd=76	n=193 $\bar{x}$=503 sd=60	n=167 $\bar{x}$=542 sd=52	N=416 $\bar{X}$=515
TOTAL		N=339 $\bar{X}$=453	N=624 $\bar{X}$=496	N=303 $\bar{X}$=536	N=1266 $\bar{X}$=494 SD=69 R=.447 η=.453

INTERNAL CONTROL:
ADDITIVE VS. COMPOSITE PREDICTION

		TIME 2 1	TIME 2 2	TIME 2 3	TOTAL
TIME 1	1	n=118 $\bar{x}$=151 sd=22	n=180 $\bar{x}$=165 sd=18	n=26 $\bar{x}$=182 sd=16	N=324 $\bar{X}$=162
TIME 1	2	n=142 $\bar{x}$=156 sd=19	n=470 $\bar{x}$=171 sd=18	n=183 $\bar{x}$=185 sd=15	N=795 $\bar{X}$=171
TIME 1	3	n=14 $\bar{x}$=170 sd=18	n=93 $\bar{x}$=174 sd=21	n=107 $\bar{x}$=189 sd=14	N=214 $\bar{X}$=182
TOTAL		N=274 $\bar{X}$=155	N=743 $\bar{X}$=170	N=316 $\bar{X}$=186	N=1333 $\bar{X}$=171 SD=21 R=.511 η=.515

NEGATIVE SCHOOL ATTITUDES:
ADDITIVE VS. COMPOSITE PREDICTION

		TIME 2			
		1	2	3	TOTAL
TIME 1	1	n=194 x=150 sd=38	n=168 x=173 sd=42	n=24 x=234 sd=61	N=386 X=165
	2	n=166 $\bar{x}$=162 sd=43	n=386 $\bar{x}$=190 sd=45	n=119 $\bar{x}$=226 sd=56	N=671 $\bar{X}$=189
	3	n=23 $\bar{x}$=175 sd=43	n=135 $\bar{x}$=211 sd=46	n=107 $\bar{x}$=251 sd=57	N=265 $\bar{X}$=224
TOTAL		N=383 $\bar{X}$=156	N=689 $\bar{X}$=190	N=250 $\bar{X}$=237	N=1322 $\bar{X}$=189 SD=54 R=.538 η=.543

BIBLIOGRAPHY

Andrews, F. M., Morgan, J. N., & Sonquist, J. A. *Multiple Classification Analysis.* Ann Arbor: Institute for Social Research, 1967.

Bachman, J. G., Kahn, R. L., Mednick, M. T., Davidson, T. N., & Johnston, L. D. *Youth in Transition, Volume I: Blueprint for a Longitudinal Study of Adolescent Boys.* Ann Arbor: Institute for Social Research, 1967.

Bachman, J. G. *Youth in Transition, Volume II: The Impact of Family Background and Intelligence on Tenth Grade Boys.* Ann Arbor: Institute for Social Research, 1970.

Bachman, J. G., Green, S., & Wirtanen, I. *Youth in Transition, Volume III: Dropping Out-Problem or Symptom?* Ann Arbor: Institute for Social Research, 1971

Bereiter, C. Some Persisting Dilemmas in the Measurement of Change. In *Problems in Measuring Change,* edited by Chester W. Harris. Milwaukee: University of Wisconsin Press, 1967.

Blalock, H. M., Jr. The Measurement Problem: A Gap Between the Languages of Theory and Research. *Methodology in Social Research,* edited by Hubert M. Blalock, Jr., and Ann B. Blalock. New York: McGraw-Hill, 1968a.

Blalock, H. M., Jr. Theory Building and Causal Inferences. *Methodology in Social Research*, edited by Hubert M. Blalock, Jr., and Ann B. Blalock. New York: McGraw-Hill, 1968b.

Blalock, H. M., Jr. Estimating Measurement Error Using Multiple Indicators and Several Points in Time. *American Sociological Review*, 1970, 35 (1), 101-111.

Bohrnstedt, G. W. A Quick Method for Determining the Reliability and Validity of Multiple-Item Scales. *American Sociological Review*, 1969, 34 (4), 542-548.

Borgatta, E. F., ed. *Sociological Methodology*. San Francisco: Jossey-Bass Inc., 1969.

Campbell, D. T. & Clayton, K. N. Avoiding Regression Effects in Panel Studies of Communication Impact. *Studies in Public Communication*, 1961, 3 (Summer), 99-118.

Campbell, D. T. Treatment-Effect Correlations and Temporal Attenuation of Relationships in Longitudinal Studies. Mimeographed. 1969.

Coleman, J. S. *Introduction to Mathematical Sociology*. London: Collier-Macmillan Limited, 1964a.

Coleman, J. S. *Models of Change and Response Uncertainty*. New Jersey: Prentice-Hall, 1964b.

Coleman, J. S. The Mathematical Study of Change. In *Methodology in Social Research*, edited by Hubert M. Blalock, Jr., and Ann B. Blalock. New York: McGraw-Hill, 1968.

Converse, P. E. Attitudes and Non-Attitudes: Continuation of a Dialogue. Revised version of a paper read at the Seventeenth International Congress of Psychology, August, 1963, in Washington, D.C.

Cronbach, L. J. Coefficient Alpha and the Internal Structure of Tests. *Psychometrika*, 1951, 16, 297-354.

Cronbach, L. J. & Furby, L. How Should We Measure Change -- or Should We? *Psychological Bulletin*, 1970, 74 (1), 68-80.

Errata. *Psychological Bulletin*, 1970, 74 (3), 218.

Glaser, B. G. & Strauss, A. L. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago: Aldine Publishing Company, 1967.

Harris, C. W., ed. *Problems in Measuring Change*. Milwaukee: University of Wisconsin Press, 1967.

Hays, W. L. *Statistics for Psychologists*. New York: Rinehart and Winston, 1963.

Heise, D. R. Separating Reliability and Stability in Test-Retest Correlation. *American Sociological Review*, 1969, 34 (February), 93-101.

Linn, R. L. & Werts, C. E. Assumptions in Making Causal Inferences from Part Correlations, Partial Correlations, and Partial Regression Coefficients. *Psychological Bulletin*, 1969, 72 (5), 307-310.

Lord, F. M. Elementary Models for Measuring Change. In *Problems in Measuring Change*, edited by Chester W. Harris. Milwaukee: University of Wisconsin Press, 1967.

McDill, E. L. & Coleman, J. S. High School Social Status, College Plans, and Interest in Academic Achievement: A Panel Analysis. *American Sociological Review*, 1963, 28 (6), 905-918.

Meehl, P. E. Nuisance Variables and the Ex Post Facto Design. *Reports from the Research Laboratories of the Department of Psychiatry, University of Minnesota*, Report No. PR-69-4, April 15, 1969.

Nunnally, J. C. *Psychometric Theory*. New York: McGraw-Hill, 1967.

Rozelle, R. M. & Campbell, D. T. More Plausible Rival Hypotheses in the Cross-Lagged Panel Correlation Technique. *Psychological Bulletin*, 1969, 71 (1), 74-80.

Shaycoft, M. *Project TALENT: The High School Years: Growth in Cognitive Skills*. Pittsburgh: American Institute for Research and School of Education, 1967.

Siegel, P. M. & Hodge, R. W. A Causal Approach to the Study of Measurement Error. In *Methodology in Social Research*, edited by Hubert M. Blalock, Jr., and Ann B. Blalock. New York: McGraw-Hill, 1968.

Sonquist, J. A. Finding Variables that Work. *The Public Opinion Quarterly*, 1969, 33, 83-95.

Sonquist, J. A. *Multivariate Model Building: The Validation of a Search Strategy*. Ann Arbor: Institute for Social Research, 1970.

Traub, R. E. A Note on the Reliability of Residual Change Scores. *Journal of Educational Measurement*, 1967, 4 (4), 253-256.

Trent, J. & Medsker, L. L. *Beyond High School*. San Francisco: Jossey-Bass Inc., 1968.

Tucker, L., Damarin, F., & Messick, S. A Base-Free Measure of Change. *Psychometrika*, 1966, 31 (4), 457-473.

Wiley, D. E. & Wiley, J. A. The Estimation of Measurement Error in Panel Data. *American Sociological Review*, 1970, 35 (1), 112-117.

INDEX

The notation "t" indicates a table or figure in which the index entry appears.